

slyX protein [Imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 AE3538
 RESULT 9

A: Accession: AF027755
A: Reference number: A82515; MUID:20365717
A: Note: for a complete list of authors see reference number A59328 below
A: Accession: D82775
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-74 <SIM>
A: Cross-references: GB:AE003912; GB:AE003849; PIDN:AAF8349
A: Experimental source: strain 9a5c
A: Accession: G82547

A;Note: the authors translated the codon ATT for residue 44 as Thr
C;Comment: This protein is male-specific.

C;Genetics:

A;Gene: anp

A;Introns: 32/3

C;Keywords: antibacterial

F;1-23/Domain: signal sequence #status predicted <SIG>

F;24-57/Product: andropin #status predicted <MAT>

Query Match 10.4%; Score 42.5; DB 2; Length 57;
Best Local Similarity 36.8%; Pred. NO. 6.2e+02;
Matches 14; Conservative 5; Mismatches 12; Indels 7; Gaps 2;

QY 22 AVSKAPQ---LVPKLDEVYNAAYNAAD---HAAPEDK 52

Db 15 AISVGPSDAVFIDILDKVENAIHNAAQVIGIGFARPEK 52

Search completed: July 11, 2002, 10:52:09

Job time: 82 sec

Db 10 LAIAGSERTQIEPVFSEKTSVYN-----YEAIVLN-----GFPE 44

```
RESULT 2
UCRX_YEAST
ID UCRX_YEAST STANDARD; PRT; 76 AA.
AC P37299;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ubiquinol-cytochrome c reductase complex 8.5 kDa protein (EC 1.10.2.2)
DE (Complex III polypeptide X).
GN Ocr10 OR YHR001BW
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=W303 YPH500;
RX MEDLINE=94230378; PubMed=8175712;
RA Brandt U., Uribe S., Schaeffer H., Trumpower B.L.;
RT "Isolation and characterization of Ocr10, the nuclear gene encoding
RT the 8.5-kDa subunit 10 of the Saccharomyces cerevisiae cytochrome bcl
RT complex."
RL J. Biol. Chem. 269:12947-12953(1994).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. OCR10 IS REQUIRED
CC FOR STABLE ASSOCIATION OF THE IRON-SULFUR PROTEIN WITH THE
CC COMPLEX.
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 ferricytochrome c = Q + 2
CC ferrocycytochrome c.
CC -!- SUBUNIT: FUNGI BCL COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY
CC SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial inner membrane.
CC -!- SIMILARITY: TO MAMMALIAN SUBUNIT XI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U07275; AAA19322.1; -
DR PIR: S48534; S48534.
DR PIR: A53768; A53768.
DR SGD: S0003529; QCR10.
KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transmembrane.
FT INIT_MET 0
FT TRANSMEM 24 44
FT SEQUENCE 76 AA; 8462 MW; E6A01B4F8EF28261 CRC64;
Query Match 10.58; Score 43; DB 1; Length 76;
Best Local Similarity 39.38; Pred. No. 3e+02;
Matches 11; Conservative 1; Mismatches 6; Indels 10; Gaps 1;
Qy 25 KAPOLVPKLDVYNAAYNAADHAAPDK 52
-| | | | |
Db 57 KIPLLGPTLE-----DHTPPEDK 74
-| | | | |
RESULT 3
ANDP_DROME
ID ANDP_DROME STANDARD; PRT; 57 AA.
AC P21663; G9VA90;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
```

```
DE Andropin precursor.
GN ANP OR ANR OR CG1361.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=91114699; PubMed=1899226;
RA Samakovlis C., Klysten P., Kimbrell D.A., Engstroem A., Hultmark D.;
RT "The andropin gene and its product, a male-specific antibacterial
RT peptide in Drosophila melanogaster.";
RL EMBO J. 10:163-169(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=B009, B141, Z10, Z18, AND Z24;
RX MEDLINE=97476321; PubMed=9335607;
RA Clark A.G., Wang L.;
RT "Molecular population genetics of Drosophila immune system genes.";
RL Genetics 147:713-724(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Ball R.J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Barille R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busam D.A., Butler H., Cadien E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck C.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ithegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Moulton G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Stappleton M., Smith T.,
RA Spier E., Spradling A.C., Stappleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Turner R., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MALE-SPECIFIC PEPTIDE WITH MODERATE ACTIVITY AGAINST
CC GRAM-POSITIVE BACTERIA.
CC -!- TISSUE SPECIFICITY: EJACULATORY DUCT OF ADULT MALES.
CC -!- INDUCTION: IN RESPONSE TO MATING.
CC -!- SIMILARITY: STRUCTURALLY RELATED TO CECROPINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 11, 2002, 10:51:47 ; Search time 11.72 seconds

(without alignments)
264.297 Million cell updates/sec

Title: US-09-696-169A-18

Perfect score: 409

Sequence: 1 ADVKTFEAAFTVSSKRRLA.....SEALHIITGPEVHAVKPGA 80

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 8305

Minimum DB seq length: 0

Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.5	10.6	71	VTAL_XENLA	P19009 xenopus lae
2	43	10.5	76	UCRX_YEAST	P37299 saccharomyc
3	42.5	10.4	57	ANDP_DROME	P21663 drosophila
4	42.5	10.4	77	R320_LACLA	O96u5 lactococcus
5	42.5	10.4	77	YC34_CYAPA	P48274 cyanophora
6	41	10.0	64	RL29_SYNPF	O24697 synechococc
7	40.5	9.9	63	YBXH_BACSU	P54426 bacillus su
8	40	9.8	36	PYY_MYOSC	P09641 myoxocephal
9	40	9.8	36	PYY_RAJRH	P29206 raja rhina
10	39.5	9.7	48	OSTC_DRONO	P15504 dromaius no
11	39	9.5	36	PYY_ORENI	P81028 oreochromis
12	39	9.5	72	YEJL_HAEIN	P44897 haemophilus
13	39	9.5	78	KEAL_ECOLI	Q52278 escherichia
14	38.5	9.4	49	OSTC_FELCA	P02821 felis silve
15	38	9.3	56	VG36_BPMO2	O22001 mycobacteri
16	38	9.3	69	YFXH_AZOC	P28486 azorhizobiu
17	38	9.3	80	Y9K0_BPP22	P57018 bacterioph
18	38	9.3	80	YDMA_SCHPO	O13915 schizosacch
19	37.5	9.2	49	OSTC_CANFA	P81455 canis fami
20	37.5	9.2	73	Y055_NPVAC	P41459 autographa
21	37	9.0	51	MLEV_MOUSE	P09542 mus musculu
22	37	9.0	69	PYY_LOPAM	P09475 lophius ame
23	37	9.0	71	ARGD_RHOCA	P30900 rhodobacter
24	37	9.0	71	CERD_CERCA	O17513 ceratitidis c
25	37	9.0	72	VG18_BPMU	Q38625 bacterioph
26	37	9.0	74	IPB2_YEAST	P01095 saccharomyc
27	37	9.0	80	CKO1_CONPL	Q9xyz0 conus pulic
28	36.5	8.9	69	Y059_NPVAC	P41463 autographa
29	36.5	8.9	70	RECX_THIPE	P37866 thiobacilla
30	36.5	8.9	72	YGEI_ECOLI	Q46789 escherichia
31	36.5	8.9	73	RL7_STAUV	P48860 staphylococ
32	36	8.8	57	Y058_NPVAC	P41462 autographa
33	36	8.8	59	YD07_METTH	O27362 methanobact

34 36 8.8 64 1 BCNB_CLOPE P15936 clostridium
35 36 8.8 66 1 CSPL_LACPL P71478 lactobacill
36 36 8.8 71 1 RL29_RICCN Q92gx4 rickettsia
37 36 8.8 71 1 RL29_RICPR Q9zcr3 rickettsia
38 36 8.8 73 1 TRM1_PYRWO P20300 pyrococcus
39 36 8.8 73 1 Y43_BPRI8 P18243 bacterioph
40 36 8.8 75 1 RL29_UREPA Q9pgq2 ureaplasma
41 36 8.8 79 1 RL22_CLOPP Q46229 clover prol
42 35.5 8.7 59 1 RL32_THETH P80339 thermus aqu
43 35.5 8.7 74 1 DR39_PEA Q01784 plasm sativ
44 35 8.6 61 1 OM06_YEAST P33448 saccharomyc
45 35 8.6 62 1 DICB_ECOLI P09557 escherichia

ALIGNMENTS

```
RESULT 1
VTAL_XENLA
ID VTAL_XENLA STANDARD; PRT; 71 AA.
AC PI9009;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Vitellogenin A1 precursor (VTG A1) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85062835; PubMed=6504704;
RA Germond J.-E., Walker P., ten Heggeler B., Brown-Luedi M., de Bony E.,
RA Wahli W.;
RT "Evolution of vitellogenin genes: comparative analysis of the
RT nucleotide sequences downstream of the transcription initiation site
RT of four Xenopus laevis and one chicken gene.";
RL Nucleic Acids Res. 12:18595-8609(1984).
CC -!- FUNCTION: PRECURSOR OF THE MAJOR EGG-YOLK PROTEINS THAT ARE
CC SOURCES OF NUTRIENTS DURING EARLY DEVELOPMENT OF OVIPAROUS
CC ORGANISMS.
CC -!- TISSUE SPECIFICITY: PRODUCED BY THE LIVER, SECRETED INTO THE BLOOD
CC AND THEN SEQUESTERED BY RECEPTOR MEDIATED ENDOCYTOSIS INTO GROWING
CC OOCYTES, WHERE IT IS GENERALLY CLEAVED, GIVING RISE TO THE
CC RESPECTIVE YOLK COMPONENTS.
CC -!- INDUCTION: By steroids (estrogen).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenseseis@sib.ch).
CC -----
CC EMBL; X01168; CAA25615.1; -.
CC PIR; A23876; A23876.
CC InterPro; IPR001747; Vitellogenin_N.
CC Pfam; PF01347; Vitellogenin_N; 1.
CC Storage protein; Phosphorylation; Signal.
CC SIGNAL 1 15
CC CHAIN 16 >71 VITELLOGENIN A1.
CC NON_TER 71 71
CC SEQUENCE 71 AA; 7827 MW; C45881649FF8A81A CRC64;
SQ
Query Match 10.6%; Score 43.5; DB 1; Length 71;
Best Local Similarity 29.6%; Pred. No. 2.5e+02;
Matches 16; Conservative 5; Mismatches 14; Indels 19; Gaps 2;
QY 19 LADAVSKAPQLVFKLDEVYNAAYNAADHAAPDKYEAFVLFHSEALHIITGPE 72
||| :|||
```


[illegible]

```

DR Pfam: PF00594; gla; 1..
DR PRINTS: PR00002; GLABONE.
DR SMART: SM0069; GLA; 1.
DR PROSITE: PS00011; GUU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxylglutamic acid; Vitamin K; Bone.
FT MOD_RES 16 16 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 20 20 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 23 23 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 22 28 BY SIMILARITY.
SQ SEQUENCE 48 AA; 5292 MW; 50A00F3BFABC7FFD CRC64;

Query Match          9.7%; Score 39.5; DB 1; Length 48;
Best Local Similarity 27.5%; Pred. No. 4.4e+02;
Matches 14; Conservative 7; Mismatches 19; Indels 11; Gaps 0

QY 10 AFTVSSKRNLA---DAVSKAPOLPKLDEVYNAAYNAADHAAPDKYEAF 56
      :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 SFVAGSSYGAAPDPLEAQREVCNELPDCDEL-----ADHGFQAYRRF 44

RESULT 11
PYL_ORENI
ID PYL_ORENI STANDARD; PRT; 36 AA.
AC P81028;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptide YY-like (PYY).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;
OC Cichlidae; Oreochromis.
OX NCBI_TaxID=8128;
RN [1]
RP SEQUENCE.
RX MEDLINE=95384941; PubMed=7656183;
RA Nguyen T.M., Wright J.R. Jr., Nielsen P.F., Conlon J.M.;
RT "Characterization of the pancreatic hormones from the Brockmann body
RT of the tilapia: implications for islet xenograft studies.";
RL Comp. Biochem. Physiol. 111C:33-44(1995).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SUBCELLULAR LOCATION: BELONGS TO THE NPY / PYY / PYY FAMILY.
DR HSPSP; P01303; IRON.
DR InterPro: IPR001955; Pancreatic_hormn.
DR Pfam: PF00159; hormone3; 1.
DR PRODOM: PD001267; Pancreatic_hormn; 1.
DR PRINTS: PR00278; PANCHORMONE.
DR SMART: SM00309; PAH; 1.
DR PROSITE: PS00265; PANCREATIC_HORMONE_1; 1.
DR PROSITE: PS02076; PANCREATIC_HORMONE_2; 1.
KW Hormone; Amidation.
FT MOD_RES 36 36 AMIDATION (BY SIMILARITY).
SQ SEQUENCE 36 AA; 4201 MW; 0246CFBC6243D7AA CRC64;

Query Match          9.5%; Score 39; DB 1; Length 36;
Best Local Similarity 56.2%; Pred. No. 3.6e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 0

QY 47 AAPED--KYEAFVLHF 60
      :||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 ASPEDWAKYHAARVRY 27

RESULT 12
YEJL_HAEIN
ID YEJL_HAEIN STANDARD; PRT; 72 AA.
AC P44897;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```
DE Hypothetical protein HI0840.
GN HI0840.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- SIMILARITY: STRONG, TO E.COLI YEJL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32766; AAC22499.1; -.
DR TIGR: HI0840;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 72 AA; 7684 MW; 7220CB06FCC35BC2 CRC64;

Query Match          9.5%; Score 39; DB 1; Length 72;
Best Local Similarity 30.0%; Pred. No. 7.9e+02;
Matches 12; Conservative 6; Mismatches 14; Indels 8; Gaps 1;

QY 38 NAAYNAADHAAPDKYFAFVLHFSEALHIAGTPEYHAVK 77
   | | | | | | | | | | | | | | | | | | | |
DB 39 NNASNLTTTSVPQTCEALQAQAFNSL-----INAYK 70

RESULT 13
KEAL_ECOLI
ID KEAL_ECOLI          STANDARD;          PRT;          78 AA.
AC Q52278;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE KLEA protein (KCRAL protein).
GN KLEA OR KCRAL.
OS Escherichia coli.
OG Plasmid IncP-beta R751.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / C600;
RX MEDLINE=95291464; PubMed=7773415;
RA Thomas C.M., Smith C.A., Ibbotson J.P., Johnston L., Wang N.;
RT "Evolution of the korA-oriv segment of promiscuous IncP plasmids."
RL Microbiology 141:1201-1210(1995).
CC -!- SIMILARITY: TO KLEC (KCRB1).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U67194; AAC64427.1; -.
KW Plasmid.
SQ SEQUENCE 78 AA; 8642 MW; 9E82DA15BA6A574E CRC64;

Query Match          9.5%; Score 39; DB 1; Length 78;
Best Local Similarity 25.8%; Pred. No. 8.6e+02;
Matches 11; Conservative 10; Mismatches 22; Indels 0; Gaps 0;

QY 2 DKYKTEAAFTSVSKRLADAVSKAPQLVPKLDDEVYNAAYNAA 44
   | | | | | | | | | | | | | | | | | | | |
DB 11 DALPNVEATDFQARRDOIEATMAEAAELVKQAEELRGKAYFAA 53

RESULT 14
OSTC_FELCA
ID OSTC_FELCA          STANDARD;          PRT;          49 AA.
AC P02821;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
DE protein) (BGP).
GN BGLAP.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RX MEDLINE=85054706; PubMed=6334077;
RA Shimomura H., Kanai Y., Sanada K.;
RT "Primary structure of cat osteocalcin."
RL J. Biochem. 96:405-411(1984).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR PIR: A03304; GECT.
DR InterPro: IPR002384; GLA_bone.
DR InterPro: IPR000294; VitK_dep_GLA.
DR Pfam: PF00594; gla; 1.
DR PRINTS: PR00002; GLABONE.
DR SMART: SM00069; GLA; 1.
DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES          9      9      HYDROXYLATION.
FT MOD_RES          17     17     GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES          21     21     GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES          24     24     GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID          23     29
SQ SEQUENCE 49 AA; 5495 MW; 93D2131FA9F656D3 CRC64;

Query Match          9.4%; Score 38.5; DB 1; Length 49;
Best Local Similarity 34.5%; Pred. No. 5.8e+02;
Matches 10; Conservative 3; Mismatches 9; Indels 7; Gaps 1;

QY 28 QLYPKLDEVYNAAYNAADHAAPDKYFAF 56
   | | | | | | | | | | | | | | | | | | | |
DB 24 ELAPDCDEL-----ADHIGFQDAYRRF 45

RESULT 15
```

```

VG36_BPMVD2'
ID VG36_BPMVD2          STANDARD;           PRT;         56 AA.
AC   O22001;
DT    15-DEC-1998 (Rel. 37, Created)
DT    15-DEC-1998 (Rel. 37, Last sequence update)
DT    15-DEC-1998 (Rel. 37, Last annotation update)
DE     Gene 36 protein (GP36).
GN      36 OR G2.
OS     Mycobacteriophage D29.
OC     Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
NCBI_TaxID=28369;
RN [1]
RP SEQUENCE FROM N.A.
MEDLINE=97419508; PubMed=9274023;
RA Ribeiro G., Viveiros M., David H.L., Costa J.V.;
R "Mycobacteriophage D29 contains an integration system similar to that
RT of the temperate mycobacteriophage L5.";
RL Microbiology 143:2701-2708(1997).
[2]
RP SEQUENCE FROM N.A.
MEDLINE=98300335; PubMed=9636706;
RA Ford M.E., Sarkis G.J., Belanger A.E., Hatfull G.F.;
RT "Genome structure of mycobacteriophage D29: implications for phage
RL evolution.";
J Mol Biol. 279:143-164(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to licensed@lsb-sib.ch).
CC -----
CC ENBL; U81553; AAB69100.1; -.
DR EMBL; AF022214; AAC18477.1; -.
SQ SEQUENCE 56 AA; 6322 MW; 62C0A18814291847 CRC64;
Query Match 9.3%; Score 38; DB 1; Length 56;
Best Local Similarity 36.7%; Pred. No. 7.7e+02;
Matches 11; Conservative 5; Mismatches 14; Indels 0; Gaps 0;
QY 7 FEAFTVSSKRNADAVSKAPQLVPKLDEV 36
| | : | : | | | : | : |
Db 13 FLGVSTIVTRYIADGRILKAVALRGPRLIIRV 42

```

Search completed: July 11, 2002, 10:55:16
Job time: 209 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

```
Run on:      July 11, 2002, 10:51:27 ; Search time 24.1 Seconds
              (without alignments)
              574.257 Million cell updates/sec
```

Title: US-09-696-169A-18
Perfect score: 409
Sequence: 1 ADKYKTFEAAFTVSKRNLA.....SEALHIIAGTPEVHAVKPGA 80

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 66515

```
Minimum DB seq length: 0
Maximum DB seq length: 80
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : STREMBL_19.*
1:  sp_archaea.*
2:  sp_bacteria.*
3:  sp_fungi.*
4:  sp_human.*
5:  sp_invertebrate.*
6:  sp_mammal.*
7:  sp_mhc.*
8:  sp_organelle.*
9:  sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_invertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*
```

SUMMARIES

Result No.	Score	%		Length	DB	ID	Description
		Query Match					
1	52	12.7	66	10	Q9LEJ4	Q9LEJ4	glycine max
2	50	12.2	66	10	Q9LEJ6	Q9LEJ6	glycine max
3	50	12.2	68	10	Q9LEJ7	Q9LEJ7	glycine max
4	50	12.2	71	10	Q9LEJ2	Q9LEJ2	glycine max
5	49.5	12.1	64	4	Q9H360	Q9H360	homo sapien
6	48.5	11.9	72	2	Q91718	Q91718	synecococc
7	48.5	11.9	79	13	Q98TP3	Q98TP3	platichtthys
8	48	11.7	68	4	Q16278	Q16278	homo sapien
9	47	11.5	61	13	Q9PWF9	Q9PWF9	coturnix co
10	46	11.2	64	2	Q46774	Q46774	escherichia
11	46	11.2	73	2	Q9YLH8	Q9YLH8	synecococc
12	46	11.2	75	10	Q9LXF0	Q9LXF0	arabidopsis
13	46	11.2	77	3	O74989	O74989	schizosacch
14	46	11.2	77	5	Q9VYS1	Q9VYS1	dtrosophila
15	46	11.2	79	16	O9CCX2	O9CCX2	mycobacteri
16	46	11.2	80	9	O9G013	O9G013	bacterioph

ALIGNMENTS

```

RESULT 1
Q9LEJ4 PRELIMINARY; PRT; 66 AA.
AC Q9LEJ4;
DT 01-OCT-2000 (T-EMBLrel. 15, Created)
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundararaman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293441; CAB96763.1; -.
FT NON_TER
SQ SEQUENCE 66 AA; 7727 MW; 6E75647455CDE94F CRC64;

Query Match 12.7%; Score 52; DB 10; Length 66;
Best Local Similarity 25.3%; Pred. No. 1.2e+02;
Matches 20; Conservative 10; Mismatches 21; Indels 28; Gaps 4;

QY 4 YKTEFAFTVSSKRNLDAYSAPOLYVKLDE-----VYNRAYNAAADHAPE-DKYEAFVL 58
||:| | |
Db 9 YKSEVIVTHV-----LPKGDHSLVAKWTFLYEKVDHTAPEPTKYKDLW 52

```

```
QY 59 HFSEALHIIAGTPEVHAVK 77
   ::::| | |
Db 53 KLTKNV-----EAHLVE 64

RESULT 2
Q9LEJ6 PRELIMINARY; PRT; 66 AA.
AC Q9LEJ6;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293437; CAB96759.1; -.
FT NON_TER 1
SQ SEQUENCE 66 AA; 7726 MW; 607B647455C7494F CRC64;

Query Match 12.2%; Score 50; DB 10; Length 66;
Best Local Similarity 29.6%; Pred. No. 2e+02;
Matches 16; Conservative 9; Mismatches 17; Indels 12; Gaps 3;

QY 29 LVPKLDE----VYNAAYNAADHAAPE-DKYEAFVLHFSEALHIIAGTPEVHAVK 77
   ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 18 VLPKGDEHSLVKWTFLYEKVDHTAPEPTKYKDLVVKLTKNV-----EAHLVE 64

RESULT 3
Q9LEJ7 PRELIMINARY; PRT; 68 AA.
AC Q9LEJ7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293437; CAB96759.1; -.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7726 MW; 607B647455C7494F CRC64;

Query Match 12.2%; Score 50; DB 10; Length 66;
Best Local Similarity 29.6%; Pred. No. 2e+02;
Matches 16; Conservative 9; Mismatches 17; Indels 12; Gaps 3;

QY 29 LVPKLDE----VYNAAYNAADHAAPE-DKYEAFVLHFSEALHIIAGTPEVHAVK 77
   ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 18 VLPKGDEHSLVKWTFLYEKVDHTAPEPTKYKDLVVKLTKNV-----EAHLVE 64

RESULT 4
Q9LEJ7 PRELIMINARY; PRT; 71 AA.
AC Q9LEJ7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. WILLIAMS 82; TISSUE=POD;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. WILLIAMS 82; TISSUE=POD;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293445; CAB96766.1; -.
FT NON_TER 1
SQ SEQUENCE 71 AA; 8318 MW; DA187778AF5D8925 CRC64;

Query Match 12.2%; Score 50; DB 10; Length 71;
Best Local Similarity 29.6%; Pred. No. 2.1e+02;
Matches 16; Conservative 9; Mismatches 17; Indels 12; Gaps 3;

QY 29 LVPKLDE----VYNAAYNAADHAAPE-DKYEAFVLHFSEALHIIAGTPEVHAVK 77
   ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 VLPKGDEHSLVKWTFLYEKVDHTAPEPTKYKDLVVKLTKNV-----EAHLVE 69

RESULT 5
Q9H360 PRELIMINARY; PRT; 64 AA.
AC Q9H360;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PRO1331.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Zhang C., Yu Y., Zhang S., Zhou G., Wei H., Bi J., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
```

```
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293436; CAB96758.1; -.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7940 MW; DA7B0455D46768CE CRC64;

Query Match 12.2%; Score 50; DB 10; Length 68;
Best Local Similarity 29.6%; Pred. No. 2e+02;
Matches 16; Conservative 9; Mismatches 17; Indels 12; Gaps 3;

QY 29 LVPKLDE----VYNAAYNAADHAAPE-DKYEAFVLHFSEALHIIAGTPEVHAVK 77
   ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 20 VLPKGDEHSLVKWTFLYEKVDHTAPEPTKYKDLVVKLTKNV-----EAHLVE 66

RESULT 4
Q9LEJ7 PRELIMINARY; PRT; 71 AA.
AC Q9LEJ7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. WILLIAMS 82; TISSUE=POD;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. WILLIAMS 82; TISSUE=POD;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293445; CAB96766.1; -.
FT NON_TER 1
SQ SEQUENCE 71 AA; 8318 MW; DA187778AF5D8925 CRC64;

Query Match 12.2%; Score 50; DB 10; Length 71;
Best Local Similarity 29.6%; Pred. No. 2.1e+02;
Matches 16; Conservative 9; Mismatches 17; Indels 12; Gaps 3;

QY 29 LVPKLDE----VYNAAYNAADHAAPE-DKYEAFVLHFSEALHIIAGTPEVHAVK 77
   ::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 23 VLPKGDEHSLVKWTFLYEKVDHTAPEPTKYKDLVVKLTKNV-----EAHLVE 69

RESULT 5
Q9H360 PRELIMINARY; PRT; 64 AA.
AC Q9H360;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PRO1331.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RA Zhang C., Yu Y., Zhang S., Zhou G., Wei H., Bi J., Dong C., Zai Y.,
RA Xu W., Gao F., Liu M., He F.;
```

```
RT "Functional prediction of the coding sequences of 11 new genes deduced
RT by analysis of cDNA clones from human fetal liver."
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF132197; AAG35543.1;
SQ SEQUENCE 64 AA; 6893 MW; 5435D402BA4D80C8 CRC64;

Query Match 12.1%; Score 49.5; DB 4; Length 64;
Best Local Similarity 28.3%; Pred. No. 2.2e+02;
Matches 13; Conservative 8; Mismatches 20; Indels 5; Gaps 1;

QY 13 VSSKRNLDADV-----SKAPQLVPKLDEVYNAAYNAADHAAPEDKY 53
DB 19 VPTAGLGDALMSSGEERGPCHKPLGQILRVHLSADSAPPQTY 64

RESULT 6
Q9L7I8 PRELIMINARY; PRT; 72 AA.
AC Q9L7I8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CPCA (FRAGMENT).
GN CPCA.
OS Synechococcus sp. P5672.
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=115734;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=PS672;
RX MEDLINE=21304427; PubMed=11411708;
RA Robertson B.R., Tezuka N., Watanabe M.M.;
RT "Phylogenetic analyses of Synechococcus strains (cyanobacteria) using
RT sequences of 16S rDNA and part of the phycocyanin operon reveal
RT multiple evolutionary lines and reflect phycobilin content."
RL Int. J. Syst. Evol. Microbiol. 51:861-871(2001).
DR EMBL: AF223433; AAF60095.1;
DR HSPF; P07122; 1CPC.
DR InterPro; IPR001659; Phycobilisome.
DR Pfam; PF00502; Phycobilisome; 1.
FT NON_TER 72
SQ SEQUENCE 72 AA; 7597 MW; 831E5FD7A9CDA95F CRC64;

Query Match 11.9%; Score 48.5; DB 2; Length 72;
Best Local Similarity 35.4%; Pred. No. 3.3e+02;
Matches 17; Conservative 5; Mismatches 13; Indels 13; Gaps 3;

QY 20 ADVSKAPQLVPKLDEVYNAAYNAADHAAPEDKYAFVLHFSEALHII 67
DB 35 ANALAAAKALAKADELVNGA-----ARP--VYNKFFPLH-----HVI 69

RESULT 7
Q98TP3 PRELIMINARY; PRT; 79 AA.
AC Q98TP3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ACETYL-COA ACETYLTTRANSFERASE (EC 2.3.1.9) (FRAGMENT).
GN ACAT1.
OS Platicthys flesus (European flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidae; Pleuronectidae; Platicthys.
OX NCBI_TaxID=8260;
RN [1]
RC SEQUENCE FROM N.A.
RP TISSUE-OVARY;
RA Williams T.D., Chipman J.K.;
```

```
RT "A DNA array to monitor the effects of environmental pollution on
RT European flounder (Platicthys flesus).";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ291983; CAC28941.1;
DR HSSP; P27796; IAFY.
KW Transferase; Acyltransferase.
FT NON_TER 1
FT NON_TER 79
SQ SEQUENCE 79 AA; 8192 MW; F491666ED133F2C02 CRC64;

Query Match 11.9%; Score 48.5; DB 13; Length 79;
Best Local Similarity 34.0%; Pred. No. 3.6e+02;
Matches 17; Conservative 5; Mismatches 25; Indels 3; Gaps 1;

QY 30 VPKLDEVY---NAAYNAADHAAPEDKYAFVLHFSEALHIIAGTPEVHAV 76
DB 7 VPKLKAFFQRENGTVAANASTLNDGAALVLTAEAAQRLNVTPLARVV 56

RESULT 8
Q16278 PRELIMINARY; PRT; 68 AA.
AC Q16278;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE VOLTAGE-DEPENDENT SODIUM CHANNEL ALPHA SUBUNIT PROTEIN
DE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=95179476; PubMed=78744451;
RA Boyle M.B., Heslip L.A.;
RT "Voltage-dependent Na+ channel mRNA expression in pregnant
RT myometrium."
RL Recept. Channels 2:249-253(1994).
DR EMBL: S75992; AAD14203.1;
DR InterPro; IPR000048; IQ.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00015; IQ; 1.
KW Ionic channel.
FT NON_TER 1
SQ SEQUENCE 68 AA; 7878 MW; E04A55B4E788BB5A CRC64;

Query Match 11.7%; Score 48; DB 4; Length 68;
Best Local Similarity 52.9%; Pred. No. 3.5e+02;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 64 LHIIAGTPEVHAVKPGA 80
DB 38 IHMIDGDRDVHATKEGA 54

RESULT 9
Q9PWF9 PRELIMINARY; PRT; 61 AA.
AC Q9PWF9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PROACTIN (FRAGMENT).
GN PRL.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
```


search completed:
Job time: 217 sec

search completed:
Job time: 217 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:50:47 ; Search time 12.98 Seconds
(without alignments)
150.543 Million cell updates/sec

Title: US-09-696-169A-18
Perfect score: 409
Sequence: 1 ADYKTFEAAFTVSSKRNL.....SEALHIIAGTPEVHAVKPGA 80

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 173753

Minimum DB seq length: 0
Maximum DB seq length: 80

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	15.9	20	1	US-08-440-861-12
2	62	15.2	20	1	US-08-440-861-11
3	59	14.4	20	1	US-08-440-861-8
4	56	13.7	20	1	US-08-440-861-13
5	54	13.2	20	1	US-08-440-861-14
6	52	12.7	20	1	US-08-440-861-10
7	50	12.2	20	1	US-08-440-861-9
8	49.5	12.1	76	3	US-08-956-307B-17
9	48.5	11.9	76	3	US-08-956-307B-18
10	48	11.7	13	1	US-08-433-854-28
11	48	11.7	13	1	US-08-174-745A-28
12	48	11.7	13	2	US-08-195-947-28
13	48	11.7	13	2	US-08-433-885-28
14	48	11.7	13	2	US-08-433-908B-28
15	48	11.7	13	4	US-08-410-614-28
16	48	11.7	20	1	US-08-440-861-15
17	43.5	10.6	36	5	PCT-US95-02885-8
18	43.5	10.6	36	5	PCT-US95-02885-8
19	43	10.5	50	4	US-09-261-855-25
20	41.5	10.1	35	2	US-08-749-526-6
21	40.5	9.9	64	2	US-08-749-526-3
22	40	9.8	12	1	US-08-433-854-20
23	40	9.8	12	1	US-08-174-745A-20
24	40	9.8	12	2	US-08-195-947-20
25	40	9.8	12	2	US-08-433-885-20
26	40	9.8	12	2	US-08-433-908B-20
27	40	9.8	12	4	US-08-410-614-20

28	40	9.8	50	4	US-08-963-851-9	Sequence 9, Appli
29	40	9.8	53	4	US-08-900-574-4	Sequence 4, Appli
30	40	9.8	61	4	US-08-936-165A-361	Sequence 361, App
31	40	9.8	63	4	US-08-900-574-3	Sequence 3, Appli
32	39.5	9.7	29	4	US-08-817-811-64	Sequence 64, Appl
33	39.5	9.7	43	1	US-08-178-477B-21	Sequence 21, Appl
34	39.5	9.7	52	1	US-08-346-849-10	Sequence 10, Appl
35	39.5	9.7	52	2	US-08-293-284A-10	Sequence 10, Appl
36	39.5	9.7	66	1	US-08-140-797-8	Sequence 8, Appli
37	39.5	9.7	66	1	US-08-486-670A-8	Sequence 8, Appli
38	39.5	9.7	66	2	US-08-749-526-2	Sequence 2, Appli
39	39	9.5	20	1	US-08-440-861-7	Sequence 7, Appli
40	38.5	9.4	60	2	US-08-968-542C-15	Sequence 15, Appl
41	38	9.3	20	3	US-08-817-926-3	Sequence 3, Appli
42	38	9.3	32	4	US-09-082-279B-1226	Sequence 1226, Ap
43	38	9.3	32	4	US-09-315-304B-1226	Sequence 1226, Ap
44	38	9.3	63	2	US-08-572-447C-9	Sequence 9, Appli
45	38	9.3	63	4	US-09-267-747-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-440-861-12
; Sequence 12, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Lugman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-12

Query Match 15.9%; Score 65; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.065;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 37 YNAYNAADHAAPEDKYAF 56
 ; | | | | | | | | | | | | | | | |
 Db 1 YDVAYKAEGATPEAKYDAF 20

RESULT 2
 US-08-440-861-11
 ; Sequence 11, Application US/08440861
 ; Patent No. 5710126
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffith, Irwin J.
 ; APPLICANT: Kuo, Mei-Chang
 ; APPLICANT: Luqman, Mohammad
 ; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
 ; TITLE OF INVENTION: ALLERGEN
 ; NUMBER OF SEQUENCES: 56
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/440,861
 ; FILING DATE: 15-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/106,016
 ; FILING DATE: 31-AUG-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amy E. Mandragouras
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 227-5941
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 20 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; US-08-440-861-11

Query Match 15.2%; Score 62; DB 1; Length 20;
 Best Local Similarity 57.9%; Pred. No. 0.15;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 27 POLVPKLDEVYNAAYNAAD 45
 ; | | | | | | | | | | | | | | | |
 Db 1 PGLIPKLDYADVAYKAAE 19

RESULT 3
 US-08-440-861-8
 ; Sequence 8, Application US/08440861
 ; Patent No. 5710126
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffith, Irwin J.
 ; APPLICANT: Kuo, Mei-Chang
 ; APPLICANT: Luqman, Mohammad
 ; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
 ; TITLE OF INVENTION: ALLERGEN
 ; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/440,861
 ; FILING DATE: 15-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/106,016
 ; FILING DATE: 31-AUG-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Amy E. Mandragouras
 ; REGISTRATION NUMBER: 36,207
 ; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617) 227-7400
 ; TELEFAX: (617) 227-5941
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 20 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FRAGMENT TYPE: internal
 ; US-08-440-861-8

Query Match 14.4%; Score 59; DB 1; Length 20;
 Best Local Similarity 75.0%; Pred. No. 0.37;
 Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 ADKYKTEFAAFTVSSK 16
 ; | | | | | | | | | | | | | | | |
 Db 5 ADKFKFEAAFSSESK 20

RESULT 4
 US-08-440-861-13
 ; Sequence 13, Application US/08440861
 ; Patent No. 5710126
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffith, Irwin J.
 ; APPLICANT: Kuo, Mei-Chang
 ; APPLICANT: Luqman, Mohammad
 ; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
 ; TITLE OF INVENTION: ALLERGEN
 ; NUMBER OF SEQUENCES: 56
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: LAHIVE & COCKFIELD
 ; STREET: 60 State Street, suite 510
 ; CITY: Boston
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/440,861
 ; FILING DATE: 15-MAY-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-13

Query Match 13.7%; Score 56; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. No. 0.86;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 47 AAPDKYEAFLHFSEALHI 66
Db 1 ATPEAKYDAFVTALRLV 20

RESULT 5
US-08-440-861-14
Sequence 14, Application US/08440861
Patent No. 5710126
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Lugman, Mohammad
TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
TITLE OF INVENTION: ALLERGEN
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,861
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-14

Query Match 13.2%; Score 54; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 1.6;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 57 VLHFSEALHIIAGTPEVHAV 76
Db 1 VTALTEALRVIAALEVHAV 20

RESULT 6
US-08-440-861-10
Sequence 10, Application US/08440861
Patent No. 5710126
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Lugman, Mohammad
TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
TITLE OF INVENTION: ALLERGEN
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,861
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-10

Query Match 12.7%; Score 52; DB 1; Length 20;
Best Local Similarity 62.5%; Pred. No. 2.8;
Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 19 LADAVSKAPQLVPKLD 34
Db 3 LATSAKAPGLIPKLD 18

RESULT 7
US-08-440-861-9
Sequence 9, Application US/08440861
Patent No. 5710126
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.


```

: APPLICANT: Kuo, Mei-Chang
: APPLICANT: Luqman, Mohammad
: TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
: TITLE OF INVENTION: ALLERGEN
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440.861
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/106,016
: FILING DATE: 31-AUG-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Amy E. Mandragouras
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 20 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: Internal
: PS-08-440-861-9

```

```

Query Match      12.2%;      Score 50; DB 1; Length 20;
Best Local Similarity 60.0%;
Matches 12; Conservative 3; Mismatches 5; Indels
QY      7 FEAFTVSSKRNLDADAVSKA 26
      1 FEAFFSSKGLLATSAAKA 20
Db

```

```

RESULT      8
US-08-956-307B-17
; Sequence 17, Application US/08956307B
; Patent No. 6090911
; GENERAL INFORMATION:
;   APPLICANT: Petka, Wendy A.
;   APPLICANT: Tirrell, David A.
;   APPLICANT: Kevin P. McGrath
;   TITLE OF INVENTION: REVERSIBLE HYDROGELS
;   NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Fish & Richardson P.C.
;   STREET: 225 Franklin Street
;   CITY: Boston
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110-2804
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows 95
;   SOFTWARE: FASTSEQ for Windows Version 2.0b
;   CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/956,307B
 FILING DATE: 22-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fasse, J. Peter
 REGISTRATION NUMBER: 32,983
 REFERENCE/DOCKET NUMBER: 07880/033001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617/542-5070
 TELEFAX: 617/542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 17:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 76 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-956-307B-17

Query Match	12.1%	Score 49.5;	DB 3;	Length 76;
Best Local Similarity	33.3%;	Pred. No. 34;		
Matches 12; Conservative	7;	Mismatches 16;	Indels 1;	Gaps 1;
Qy	17	RNLADAYSKAPQLVPKL-DEVYNAAYNAAADHAAPED	51	
Dd	35	RSLEDEAAEAEOKYSRLKNEIETDKAETGDHVAPRD	70	

```

RESULT          9
US-08-956-307B-18
; Sequence 18, Application US/08956307B
; Patent No. 6090911
;
; GENERAL INFORMATION:
; APPLICANT: Petka, Wendy A.
; APPLICANT: Tirrell, David A.
; APPLICANT: Kevin P. McGrath
; TITLE OF INVENTION: REVERSIBLE HYDROGELS
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows version 2.0.b
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956,307B
; FILING DATE: 22-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07880/033001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

```

US-08-174-745A-28

ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,947
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-195-947-28

Query Match 11.7%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 67 IAGTPEVHAVKP 78
||| |||||
DB 1 IAGALEVHAVKP 12

RESULT 13
US-08-433-885-28
; Sequence 28, Application US/08433885
; Patent No. 5869333
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,885
FILING DATE: 02-MAY-1995

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-039C2D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-885-28

Query Match 11.7%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 67 IAGTPEVHAVKP 78
||| |||||
DB 1 IAGALEVHAVKP 12

RESULT 14
US-08-433-908B-28
; Sequence 28, Application US/08433908B
; Patent No. 5965455
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,908B
FILING DATE: 02-MAY-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Mandragouras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-039C2D4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-908B-28

Query Match 11.7%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 67 IAGTPEVHAVKP 78
Db 1 IAGALEVHAVKP 12

RESULT 15

US-08-410-614-28
; Sequence 28, Application US/08410614
; Patent No. 6277383
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,614
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,947
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-410-614-28

Query Match 11.7%; Score 48; DB 4; Length 13;
Best Local Similarity 83.3%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 67 IAGTPEVHAVKP 78
Db 1 IAGALEVHAVKP 12

Search completed: July 11, 2002, 10:51:45
Job time: 58 sec

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: H82699

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-56 <SIM>

A:Cross-references: GB:AE003962; GB:AE003849; NID:g9106270; PIDN:AAF84094.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigh

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tshakoko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF1285

Query Match 14.8%; Score 44; DB 2; Length 56;

Best Local Similarity 25.7%; Pred. No. 1.5e+02;

Matches 9; Conservative 10; Mismatches 16; Indels 0; Gaps 0;

OY 8 PKLDEVYNAAYNAADHAAPEDKYAEFVLHSEALH 42

Db 11 PKIFSPHNDIHQPSMHSMHPNHQHTAQKTNVSFAMH 45

RESULT 3

YFIS

peptide YY - shorthorn sculpin

N:Alternate names: pancreatic hormone; pancreatic polypeptide

C:Species: Myoxocephalus scorpius (shorthorn sculpin, daddy sculpin)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 20-Mar-1998

A:Accession: A60309; B26993

R:Conlon, J.M.; Schmidt, W.E.; Gallwitz, B.; Falkmer, S.; Thim, L.

Regul. Pept. 16, 261-268, 1986

A:Title: Characterization of an amidated form of pancreatic polypeptide from the daddy s

A:Reference number: A60309; MUID:87176585

A:Accession: A60309

A:Molecule type: protein

A:Residues: 1-36 <CON>

A:Experimental source: islet organs (Brockmann bodies)

R:Cutfield, S.M.; Carne, A.; Cutfield, J.F.

FEBS Lett. 214, 57-61, 1987

A:Title: The amino-acid sequences of sculpin islet somatostatin-28 and peptide YY.

A:Reference number: A91376; MUID:87190954

A:Accession: B26993

A:Molecule type: protein

A:Residues: 1-36 <CUT>

A:Experimental source: islet organs (Brockmann bodies)

C:Comment: Peptide YY and pancreatic peptide are identical in this fish but are distinct

C:Comment: The protein has been demonstrated in the pyloric but not the splenic Brockman

C:Superfamily: pancreatic hormone

C:Keywords: amidated carboxyl end; hormone; pancreas

F:36/Modified site: amidated carboxyl end (Tyr) #status experimental

Query Match 13.5%; Score 40; DB 1; Length 36;

Best Local Similarity 52.9%; Pred. No. 2.8e+02;

Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

OY 23 HAAPED--KYAEFVLHF 37

Db 11 NASPDWAKYHAARVHY 27

RESULT 4

S70651

leukotriene-A4 hydrolase (EC 3.3.2.6) short isoform - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 07-May-1999

C:Accession: S70651

R:Jendraschak, E.; Kaminski, W.E.; Kiefl, R.; von Schacky, C.

Biochem. J. 314, 733-737, 1996

A:Title: The human leukotriene A(4) hydrolase gene is expressed in two alternatively-

A:Reference number: S70627; MUID:96177841

A:Accession: S70651

A:Molecule type: mRNA

A:Residues: 1-49 <JEN>

C:Cross-references: EMBL:U43411

C:Superfamily: leukotriene-A4 hydrolase

C:Keywords: alternative splicing; ether hydrolase

Query Match 13.5%; Score 40; DB 2; Length 49;

Best Local Similarity 33.3%; Pred. No. 4e+02;

Matches 10; Conservative 9; Mismatches 7; Indels 4; Gaps 2;

OY 7 VPKLDEVYNAAYNAADHAAPEDKYAEFVLH 36

Db 8 IKRMQEVYN--FNAINNS--EIRFMAALH 33

RESULT 5

S52358

hypothetical protein - human

C:Species: Homo sapiens (man)

C:Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 05-Nov-1999

C:Accession: S52358

R:Agrykstritis, A.; Leversha, M.A.; Ferguson-Smith, M.; Moschonas, M.K.

submitted to the EMBL Data Library, March 1993

A:Description: A cosmid clone mapped to human chromosome 11p15 detects a Taq I restri

A:Reference number: S52355

A:Accession: S52358

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-43 <AGR>

A:Cross-references: EMBL:X72882; NID:g667005; PIDN:CAA51394.1; PID:g667006

Query Match 13.1%; Score 39; DB 2; Length 43;

Best Local Similarity 36.7%; Pred. No. 4.6e+02;

Matches 11; Conservative 4; Mismatches 13; Indels 2; Gaps 1;

OY 24 AAPEDKYAEFVLHFSEALHIIAGTPEVHAV 53

Db 16 ASFSDKSEP--ASPSDFMHLITKKPSVEQV 43

RESULT 6

R3KM72

ribosomal protein S7-2 - Chlamydomonas reinhardtii chloroplast

C:Species: chloroplast Chlamydomonas reinhardtii

C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 22-Jun-1999

C:Accession: S11897

R:Robertson, D.; Boynton, J.E.; Gillham, N.W.

Mol. Gen. Genet. 221, 155-163, 1990

A:Title: Cotranscription of the wild-type chloroplast atpE gene encoding the CF(1)/CF

1ft mutations in atpE.

A:Reference number: S11897; MUID:90318312

A:Accession: S11897

A:Molecule type: DNA

A:Residues: 1-52 <ROB>

A:Cross-references: EMBL:X53977; NID:g11429; PIDN:CAA37927.1; PID:g11430

C:Genetics:

A:Superfamily: chloroplast

C:Superfamily: Chlamydomonas chloroplast ribosomal protein S7-2

Db 29 LDKVENAIHNAQVGIGFAKPFK 52

Db 24 ELNPDCEL-----ADHIGFDAYRRF 45

RESULT 11

S50084

single stranded DNA-binding protein H16 - green monkey (fragments)
N:Alternate names: RNA-binding protein K homolog
C:Species: Cercopithecus aethiops (green monkey, grivet)
C:Date: 23-May-1997 #sequence_revision 23-May-1997 #text_change 19-Feb-1999
C:Accession: S50084
R:Galliard, C.; Cabanuss, E.; Strauss, F.
Nucleic Acids Res. 22, 4183-4186, 1994
A:Title: Identity of the RNA-binding protein K of hnRNP particles with protein H16, a se
A:Reference number: S50084; MUID:950231188
A:Accession: S50084
A:Molecule type: protein
A:Residues: 1-16;17-28;29-41 <GAI>
A:Experimental source: cell line CV1

Query Match 12.8%; Score 38; DB 2; Length 41;
Best Local Similarity 41.2%; Pred. No. 5.7e+02;
Matches 7; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 14 YNAAYNAADHAAPEDKY 30

|||: | | | |

Db 3 YNASVXVPDXXGPEGSY 19

RESULT 12

G95302

hypothetical protein Sma0625 [imported] - Sinorhizobium meliloti (strain 1021) magaplasmi
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: G95302
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bows
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium melilot
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: G95302
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-53 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK64985.1; PID:gl4523412; GSPDB:GN00165
R:Experimental source: strain 1021, megaplasmaid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Fuhrer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
peia, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0625
A:Genome: plasmid

Query Match 12.8%; Score 38; DB 2; Length 53;
Best Local Similarity 41.2%; Pred. No. 7.7e+02;
Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Qy 32 APVLFHSEALHIIAGTP 48

|||: | | | |

Db 18 APMLFLQSVVHAFAGOP 34

RESULT 13

B41397

hypothetical protein 1 - mouse
C:Species: Mus musculus (house mouse)

C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 03-May-1994
C:Accession: B41397
R:Miller, D.A.; Lee, A.; Matsui, Y.; Chen, E.Y.; Moses, H.L.; Derynck, R.
Mol. Endocrinol. 3, 1926-1934, 1989
A:Title: Complementary DNA cloning of the murine transforming growth factor-beta3 (TG
t tissues
A:Reference number: A41397; MUID:90190650
A:Accession: B41397
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-34 <MIL>
A:Cross-references: GB:M32745

Query Match 12.5%; Score 37; DB 2; Length 34;
Best Local Similarity 37.5%; Pred. No. 6.2e+02;
Matches 6; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 36 HFSEALHIIAGTPEVH 51

||: | | : ||| |

Db 16 YFVQGLPALPGSPEAH 31

RESULT 14

A26781

peptide YG - American goosefish
C:Species: Lophius americanus (American goosefish)
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 12-Apr-1995
C:Accession: A26781
R:Andrews, P.C.; Hawke, D.; Shively, J.E.; Dixon, J.E.
Endocrinology 116, 2677-2681, 1985
A:Title: A nonamidated peptide homologous to porcine peptide YY and neuropeptide YY.
A:Reference number: A26781; MUID:85203740
A:Accession: A26781
A:Molecule type: protein
A:Residues: 1-37 <AND>
C:Superfamily: pancreatic hormone

Query Match 12.5%; Score 37; DB 2; Length 37;
Best Local Similarity 47.1%; Pred. No. 6.8e+02;
Matches 8; Conservative 4; Mismatches 3; Indels 2; Gaps 1;

Qy 23 HAAPED-KYEAFLVLFH 37

||: ||| | | | |

Db 11 NASPEDWASYQAARHY 27

RESULT 15

T36238

hypothetical protein SCE39.28 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36238
R:Olliver, K.; Harris, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21577
A:Accession: T36238
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-42 <OLI>
A:Cross-references: EMBL:AL049573; PIDN:CAB40336.1; GSPDB:GN00070; SCOEDB:SCE39.28
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SCE39.28

Query Match 12.5%; Score 37; DB 2; Length 42;
Best Local Similarity 58.3%; Pred. No. 7.8e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 46 GTPEVHAVKPGA 57

| | | | |

Thu Jul 11 11:40:43 2002

us-09-696-169a-19.closed.rpr

Page 5

Db 8 *GVPSRHAVREGA 19

Search completed: July 11, 2002, 10:56:37
Job time: 130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:56:12 ; Search time 10.18 Seconds

(without alignments)
216.799 million cell updates/sec

Title: US-09-696-169A-19

Perfect score: 297

Sequence: 1 SKAPQLVPKLDEVYNAAAYNA.....SEALHIIAGTPEVHAVKPGA 57

Scoring table: BLOSUM62

Learning table: 2000000
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 4574

Minimum DB seq length: 0

Maximum DB seq length:	0
Maximum DB seq length:	57

Post-processing: Minimum Match 0%

Loss processing. Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt 40:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	40	13.5	36	1	PVY_MYOSC	P09641 myoxocephal
2	40	13.5	36	1	PVY_RAJRH	P29206 raja rhina
3	39	13.1	36	1	PVY_ORENI	P81028 oreochromis
4	39	13.1	57	1	ANDP_DROME	P21663 drosophila
5	38.5	13.0	49	1	OSTC_FELCA	P02821 felis silve
6	37.5	12.6	49	1	OSTC_CANFA	P81455 canis famil
7	34.5	11.6	48	1	OSTC_DRONO	P15504 dromaelus no
8	34.5	11.6	49	1	OSTC_MACFA	P02819 macaca fasc
9	34.5	11.6	49	1	OSTC_RABIT	P39056 oryctolagus
10	34.5	11.6	49	1	OSTC_XENLA	P40147 xenopus lae
11	34	11.4	46	1	PSBK_SYNEL	Q9f1k9 synechococc
12	34	11.4	48	1	RK32_VICFA	P15820 vicia faba
13	34	11.4	51	1	INS_BALBO	P01314 balaenopter
14	34	11.4	51	1	INS_BALPH	P01312 balaenopter
15	34	11.4	51	1	INS_CAMDR	P01320 camelus dro
16	34	11.4	51	1	INS_CAPHI	P01319 capra hircu
17	34	11.4	51	1	INS_DIDMA	P18109 didelphis m
18	34	11.4	51	1	INS_ELEMA	P01316 elephas max
19	34	11.4	51	1	INS_FELCA	P06306 felis silve
20	33.5	11.3	51	1	YNFN_ECOLI	P76157 escherichia
21	33	11.1	51	1	INS_ANGRO	P42633 anguilla ro
22	33	11.1	56	1	CAQS_CANFA	P12336 canis famil
23	33	11.1	57	1	INS_PETMA	P14806 petromyzon
24	32.5	10.9	39	1	FUC3_RAT	P80349 rattus norv
25	32	10.8	44	1	BAB4_BABBO	P14201 babesia bov
26	32	10.8	50	1	INS_KATPE	P01340 katsuwonus
27	32	10.8	51	1	INS_ACOCA	P01324 acomys cahi
28	32	10.8	51	1	INS_ANSAN	P07454 anser anser
29	32	10.8	51	1	INS_HYSCR	P01328 hystrix cri
30	32	10.8	51	1	INS_ORNAN	Q9tcq7 ornithorhyn
31	32	10.8	51	1	INS_TRASC	P31887 trachemys s
32	32	10.8	51	1	MLEV_MOUSE	P09542 mus musculu
33	32	10.8	52	1	INS_ACIGU	P81423 acipenser g

ALIGNMENTS

```

RESULT 1
PYV_MYOSC
ID PYV_MYOSC STANDARD; PRT; 36 AA.
AC P09641;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptide YY-like (PPY).
OS Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;
OC Cottoidae; Cottidae; Myoxocephalus.
OC NCBI_Taxid=8097;
RN [1]
RN SEQUENCE.
RC TISSUE=Pancreas;
RX MEDLINE=87190954; PubMed=2883025;
RX Cutfield S.M.; Carne A.; Cuffield J.F.;
RA "The amino-acid sequences of sculpin islet somatostatin-28 and
RT peptide YY.";
RL FEBS Lett. 214:57-61(1987).
RN [2]
RN SEQUENCE.
RX MEDLINE=87176585; PubMed=3562898;
RX Conlon J.M.; Schmidt W.E.; Gallwitz B.; Falkner S.; Thim L.;
RA "Characterization of an amidated form of pancreatic polypeptide from
RT the daddy sculpin (Cottus scorpius).";
RL Regul. Pept. 16:261-268(1986).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE NPY / PPY / PPY FAMILY.
DR PIR; A60309; YFIF5.
DR HSSP; P01303; IRON.
DR InterPro; IPR001955; Pancreatic_hormn.
DR Pfam; PF00159; hormone3.1.
DR PRINTS; PR00278; PANCHORMONE.
DR PRODOM; PD001267; Pancreatc_hormn; 1.
DR SMART; SM00309; PAH; 1.
DR PROSITE; PS00265; PANCREATIC_HORMONE_1; 1.
DR PROSITE; PS0276; PANCREATIC_HORMONE_2; 1.
KW Hormone; Amidation.
FT MOD_RES 36 36
FT SEQUENCE 36 AA; 4170 MW; 16F331B73643D7AA CRC64;

Query Match 13.5%; Score 40; DB 1; Length 36;
Best Local Similarity 52.9%; Pred. No. 1e+02;
Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1

QY 23 HAAPED--KYEAFLVLFH 37
:|:|:| | | | |
Db 11 NASPDWAKYHAARHY 27

RESULT 2
PYV_RAJRH

```

KW	Hormone; Amidation.
FT	MOD_SEQUENCE 36 36 AMIDATION (BY SIMILARITY).
SQ	SEQUENCE 36 AA; 4201 MW; 0246CFBC6243D7AA CRC64;
Query Match	13.1%; Score 39; DB 1; Length 36;
Best Local Similarity	56.2%; Pred. No. 1.4e+02;
Matches	9; Conservative 2; Mismatches 3; Indels 2; Gaps
QY	24 AAPED--KYEAFLVLFH 37 I : I I I I I I I I :
Dd	12 ASPEDWKYHAARVHY 27 I : I I I I I I I I :
RESULT 4	
ID	ANDP_DROME STANDARD; PRT; 57 AA.
AC	P2163; Q9VA90;
DT	01-MAY-1991 (Rel. 18, Created)
DT	01-MAY-1991 (Rel. 18, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Andropin precursor.
GN	ANP OR ANR OR CG1361.
OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CANTON-S;
RX	MEDLINE=91114699; PubMed=1899226;
RA	Samakovlis C., Kysten P., Kimbrell D.A., Engstroem A., Hultmark D.;
RT	"The andropin gene and its product, a male-specific antibacterial
RT	peptide in Drosophila melanogaster."
RL	EMBO J. 10:163-169(1991).
[2]	
RP	SEQUENCE FROM N.A.
RC	STRAIN=B009, B141, Z10, Z18, AND Z24;
RX	MEDLINE=97476321; PubMed=9335607;
RA	Clark A.G., Wang L.;
RT	"Molecular population genetics of Drosophila immune system genes.";
RL	Genetics 147:713-724(1997).
[3]	
RP	SEQUENCE FROM N.A.
RC	STRAIN=BERKELEY.
RX	MEDLINE=20196006; PubMed=10731132;
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA	Ananatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA	George R.A., Lewis S.R., Richards S., Ashburner M., Henderson S.N.,
RA	Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA	Arkil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA	Beeoon K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA	Cherry J.M., Clawley S., Dahlke C., Davenport L.B., Davies P.,
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA	Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA	Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA	Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA	Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA	Lasko P., Lei Y., Levitsky A.C., Li J., Li Z., Liang Y., Lin X.,
RA	Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

```
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "the genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MALE-SPECIFIC PEPTIDE WITH MODERATE ACTIVITY AGAINST
CC GRAM-POSITIVE BACTERIA.
CC -!- TISSUE SPECIFICITY: EJACULATORY DUCT OF ADULT MALES.
CC -!- INDUCTION: IN RESPONSE TO MATING.
CC -!- SIMILARITY: STRUCTURALLY RELATED TO CECROPINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15672; CAA34842.1; -
DR EMBL: X56726; CAA40046.1; -
DR EMBL: AF018985; AAB82482.1; -
DR EMBL: AF018986; AAB82483.1; -
DR EMBL: AF018989; AAB82486.1; -
DR EMBL: AF018990; AAB82487.1; -
DR EMBL: AF018992; AAB82489.1; -
DR EMBL: AE003773; AAF57024.1; -
DR PIR: SI3450; SI3450.
DR FlyBase; FBgn0000094; Anp.
DR Insect Immunity; Antibiotic; Signal.
KW SIGNAL 1 23 POTENTIAL.
FT PEPTIDE 24 57 ANDROPIN.
FT SEQUENCE 57 AA: 6151 MW; EEA8CCCF59A09AB5 CRC64;
QY 10 LDEVYNAAYNAAD----HAAPEDK 29
||| ||| |||
||| ||| |||
Db 29 LDKVENAIHNAAQVGIGFAKPFKE 52
||| ||| |||
||| ||| |||
RESULT 5
OSTC_FELCA STANDARD; PRT; 49 AA.
ID AC P02821;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
DE protein) (BGP).
GN BGLAP.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RX MEDLINE=85054706; PubMed=6334077;
RA Shimomura H., Kanai Y., Sanada K.;
RT "Primary structure of cat osteocalcin.";
RL J. Biochem. 96:405-411(1984).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES 9 9 HYDROXYLATION.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29
SQ SEQUENCE 49 AA; 5495 MW; 93D2131FA9F656D3 CRC64;
```

```
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR PIR; A03304; GECT.
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES 9 9 HYDROXYLATION.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29
SQ SEQUENCE 49 AA; 5495 MW; 93D2131FA9F656D3 CRC64;

Query Match 13.0%; Score 38.5; DB 1; Length 49;
Best Local Similarity 34.5%; Pred. No. 2.2e+02;
Matches 10; Conservative 3; Mismatches 9; Indels 7; Gaps 1;

QY 5 QLVPKLDEVYNAAYNAADHAAPEDKYEAF 33
||| ||| |||
||| ||| |||
Db 24 ELNPDCEL-----ADHIGFQDAYRRF 45

RESULT 6
OSTC_CANFA STANDARD; PRT; 49 AA.
ID AC P81455;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
DE protein) (BGP).
GN BGLAP.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE.
RX MEDLINE=93318657; PubMed=8101026;
RA Colombo G., Fanti P., Yao C., Malluche H.H.;
RT "Isolation and complete amino acid sequence of osteocalcin from
RT canine bone.";
RL J. Bone Miner. Res. 8:733-743(1993).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES 9 9 HYDROXYLATION (BY SIMILARITY).
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29
SQ SEQUENCE 49 AA; 5524 MW; 43121D015817CEA6 CRC64;
```

Query Match 12.6%; Score 37.5; DB 1; Length 49;
 Best Local Similarity 31.0%; Pred. No. 3e+02;
 Matches 9; Conservative 5; Mismatches 8; Indels. 7; Gaps 1;

QY 5 QLVPKLDEVYNAAYNAADHAAPEDKYEAF 33
 :| | | | :| | | :| | |
 Db 24 ELNPNCDL-----ADHIGFQYAYRRF 45

RESULT 7

OSTC_DRONO STANDARD; PRT; 48 AA.
 AC P15504;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-protein) (BGP).
 GN BGLAP.
 OS Dromaius novaehollandiae (Emu).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Palaeognathae; Casuariformes; Dromalidae;
 OC Dromaius.
 ON NCBI_TaxID=8790;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=88134266; PubMed=3501719;
 RA Huq N.L., Tseng A., Chapman G.E.;
 RT "The amino acid sequence of Emu osteocalcin: gas phase sequencing of
 FT Gla-containing proteins.";
 RL Biochem. Int. 15:271-277(1987).
 CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
 CC STRONGLY TO APATITE AND CALCIUM.
 CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
 CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
 CC BINDING OF CALCIUM.
 CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
 CC FAMILY.
 DR PIR: S02208; S02208.
 DR InterPro: IPR002384; GLA_bone.
 DR InterPro: IPR000294; VitK_dep_GLA.
 DR Pfam: PF00594; gla; 1.
 DR PRINTS: PS00002; GLABONE.
 DR SMART: SM00069; GLA; 1.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
 KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
 KW Bone.
 FT MOD_RES 16 16 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 16 16 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 20 20 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 23 23 GAMMA-CARBOXYGLUTAMIC ACID.
 FT DISULFID 22 28 BY SIMILARITY.
 SQ SEQUENCE 48 AA; 5292 MW; 50A00F3BFABC7FFD CRC64;

Query Match 11.6%; Score 34.5; DB 1; Length 48;
 Best Local Similarity 31.0%; Pred. No. 6.8e+02;
 Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

QY 5 QLVPKLDEVYNAAYNAADHAAPEDKYEAF 33
 :| | | | :| | | :| | |
 Db 23 ELNPNCDL-----ADHIGFQYAYRRF 44

RESULT 8

OSTC_MACFA STANDARD; PRT; 49 AA.
 AC P02819;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-protein) (BGP).
 GN BGLAP.
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 ON NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92175242; PubMed=1794506;
 RA Virdi A.S., Willis A.C., Hauschka P.V., Triffitt J.T.;
 RT "Primary amino acid sequence of rabbit osteocalcin.";
 RL Biochem. Soc. Trans. 19:373S-373S(1991).
 CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
 CC STRONGLY TO APATITE AND CALCIUM.
 CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
 CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
 CC BINDING OF CALCIUM.
 CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
 CC FAMILY.
 DR PIR: A61280; A61280.

GN BGLAP.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Macaca.
 ON NCBI_TaxID=9541;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=82182842; PubMed=6978733;
 RA Hauschka P.V., Carr S.A., Biemann K.;
 RT "Primary structure of monkey osteocalcin.";
 RL Biochemistry 21:638-642(1982).
 CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
 CC STRONGLY TO APATITE AND CALCIUM.
 CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
 CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
 CC BINDING OF CALCIUM.
 CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
 CC FAMILY.
 DR PIR: A03302; GEMKI.
 DR InterPro: IPR002384; GLA_bone.
 DR InterPro: IPR000294; VitK_dep_GLA.
 DR Pfam: PF00594; gla; 1.
 DR PRINTS: PR00002; GLABONE.
 DR SMART: SM00069; GLA; 1.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
 KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
 KW Bone.
 FT MOD_RES 9 9 HYDROXYLATION.
 FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
 FT DISULFID 23 29 BY SIMILARITY.
 SQ SEQUENCE 49 AA; 5743 MW; C20116014D0C4958 CRC64;

Query Match 11.6%; Score 34.5; DB 1; Length 49;
 Best Local Similarity 31.0%; Pred. No. 7e+02;
 Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

QY 5 QLVPKLDEVYNAAYNAADHAAPEDKYEAF 33
 :| | | | :| | | :| | |
 Db 24 ELNPNCDL-----ADHIGFQYAYRRF 45

RESULT 9

OSTC_RABIT STANDARD; PRT; 49 AA.
 AC P39056;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein)
 DE (Bone Gla-protein) (BGP).
 GN BGLAP.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 ON NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=92175242; PubMed=1794506;
 RA Virdi A.S., Willis A.C., Hauschka P.V., Triffitt J.T.;
 RT "Primary amino acid sequence of rabbit osteocalcin.";
 RL Biochem. Soc. Trans. 19:373S-373S(1991).
 CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
 CC STRONGLY TO APATITE AND CALCIUM.
 CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
 CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
 CC BINDING OF CALCIUM.
 CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
 CC FAMILY.
 DR PIR: A61280; A61280.

```
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW BONE.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 9 9 HYDROXYLATION.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29 BY SIMILARITY.
SQ SEQUENCE 49 AA; 5431 MW; 7B218871F0312253 CRC64;

Query Match 11.6%; Score 34.5; DB 1; Length 49;
Best Local Similarity 31.0%; Pred. No. 7e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

Qy 5 QLVPKLDEVYNAAYNAADHAAPEDKYEF 33
:| | | | | | | | | | | | | | |
Db 24 ELNPDCEL-----ADQVGLQDAYQRF 45

RESULT 10
ID OSTC_XENLA STANDARD; PRT; 49 AA.
AC P40147;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
protein) (BGP).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE.
RX MEDLINE=96139691; PubMed=8567186;
RA Canceda M., Williamson M.K., Price P.A.;
RT "Amino-acid sequence of bone Gla protein from the African clawed toad
xenopus laevis and the fish Sparus aurata.";
RL Int. J. Pept. Protein Res. 46:419-423(1995).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU-CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Bone.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29 BY SIMILARITY.
SQ SEQUENCE 49 AA; 5360 MW; 7A9A8F63A12E6047 CRC64;

Query Match 11.6%; Score 34.5; DB 1; Length 49;
Best Local Similarity 31.0%; Pred. No. 7e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

Qy 5 QLVPKLDEVYNAAYNAADHAAPEDKYEF 33
```

```
Db 24 ELNPDCEL-----ADHIGFOEAYRRF 45

RESULT 11
PSBK_SYNEL
ID PSBK_SYNEL STANDARD; PRT; 46 AA.
AC Q9FLK9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
PSBK.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RA Katoh H., Ikeuchi M.;
RT "Cloning and disruption of the psbK gene from thermophilic
Thermosynechococcus elongatus.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
CC OF PHOTOSYSTEM II.
CC -!- SIMILARITY: BELONGS TO THE PSBK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB052850; BAB20628.1; --
DR InterPro; IPR003687; Psbk.
DR Pfam; PF02533; Psbk; 1.
KW Photosystem II.
FT PROPEP 1 9 BY SIMILARITY.
FT CHAIN 10 46 PHOTOSYSTEM II REACTION CENTER PROTEIN K.
SQ SEQUENCE 46 AA; 5026 MW; 45C5197F4B50E398 CRC64;

Query Match 11.4%; Score 34; DB 1; Length 46;
Best Local Similarity 40.0%; Pred. No. 7.5e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 39 EALHIITGTPVHAV 53
:| | | | | | | | | | | | | | |
Db 3 DALVLVAKLPEAYAI 17

RESULT 12
RK32_VICFA
ID RK32_VICFA STANDARD; PRT; 48 AA.
AC P15820;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L32.
GN RPL32.
OS Vicia faba (Broad bean).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Vicia.
OX NCBI_TaxID=3906;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206803; PubMed=2320425;
RA Herdenberger F., Pillay D.T.N., Steinmetz A.;
RT "Sequence of the trnH gene and the inverted repeat structure deletion
site of the broad bean chloroplast genome.";
```

```
RL Nucleic Acids Res. 18:1297-1297(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89208884; PubMed=3242868;
RA Herdenberger F., Weil J.H., Steinmetz A.;
RT "Organization and nucleotide sequence of the broad bean chloroplast
genes trnI-UAG, ndhF and two unidentified open reading frames.";
RL Curr. Genet. 14:609-615(1988).
CC -!- SIMILARITY: BELONGS TO THE L32P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51471; CAA35833.1; -.
DR PIR: S08495; R5VF32.
DR PIR: S08444; S08444.
DR Mendel; 5435; VICfa:rpl32; 1.
DR InterPro: IPR002677; Ribosomal_L32p.
DR Pfam: PF01783; Ribosomal_L32p; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 48 AA; 5407 MW; 146372D6C9BC34DB CRC64;

Query Match 11.4%; Score 34; DB 1; Length 48;
Best Local Similarity 42.9%; Pred. No. 7.9e+02;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 30 YEAFVLHFSEALHIIAGTPEV 50
I: I : I I : I : I :
D 25 YKALKAFSLADSLTGTSKV 45

RESULT 13
INS_BALBO
ID INS_BALBO STANDARD; PRT; 51 AA.
AC P01314.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE INSULIN.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
RN [1]
RP SEQUENCE.
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin.";
RL Nature 181:1468-1469(1958).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A01582; INWH1.
DR HSSP: P01315; 9INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
DR
```

```
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 11.4%; Score 34; DB 1; Length 51;
Best Local Similarity 32.0%; Pred. No. 8.4e+02;
Matches 8; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

QY 36 HFSEALHIIAG-----TPEVHAVK 54
I : I I : I : I :
D 10 HLVEALYLCVGERGFFVTPKAGIVE 34

RESULT 14
INS_BALPH
ID INS_BALPH STANDARD; PRT; 51 AA.
AC P01312.
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE INSULIN.
GN INS.
OS Balaenoptera physalus (Finback whale) (Common rorqual), and
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9770, 9755;
RN [1]
RP PARTIAL SEQUENCE.
RC SPECIES=B.Physalus;
RA Hama H., Titani K., Sakaki S., Narita K.;
RT "The amino acid sequence in fin-whale insulin.";
RL J. Biochem. 56:285-293(1964).
RN [2]
RP SEQUENCE.
RC SPECIES=P.catodon;
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin.";
RL Nature 181:1468-1469(1958).
RN [3]
RP SEQUENCE.
RC SPECIES=P.catodon;
RA Harris J.I., Sanger F., Naughton M.A.;
RT "Species differences in insulin.";
RL Arch. Biochem. Biophys. 65:427-438(1956).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR: A91918; INWHF.
DR PIR: A93142; INWHP.
DR HSSP: P01315; 9INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam: PF00049; Insulin; 1.
DR PRINTS: PR00276; INSULIN.
DR PRINTS: PR00277; INSULINB.
DR SMART: SM00078; ILGF; 1.
DR PROSITE: PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
```

```

Query Match      11.4%; Score 34; DB 1; Length 51;
Best Local Similarity 32.0%; Pred. No. 8.4e+02;
Matches 8; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

QY 36 HFSEALHIIAG-----TPEVHAVK 54
Db 10 HLVEALYLVCGERGFYFYPKAGIVE 34

```

Search completed: July 11, 2002, 10:59:38
Job time: 206 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:55:52 ; Search time 24.05 seconds
(without alignments)
410.009 Million cell updates/sec

Title: US-09-696-169a-19

Perfect score: 297

Sequence: 1 SKAPQLVPKLDVYNAAYNA.....SEALHIIAGTPEVHAVKPGA 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 37469

Minimum DB seq length: 0

Maximum DB seq length: 57

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:**

1: sp_archaea:**

2: sp_bacteria:**

3: sp_fungi:**

4: sp_human:**

5: sp_invertebrate:**

6: sp_mammal:**

7: sp_mhc:**

8: sp_organelle:**

9: sp_phage:**

10: sp_plant:**

11: sp_rodent:**

12: sp_virus:**

13: sp_vertebrate:**

14: sp_unclassified:**

15: sp_virus:**

16: sp_bacteriap:**

17: sp_archaeap:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45.5	15.3	54	16 Q9PA18	Q9pai8 xylella fas
2	44	14.8	42	10 Q9LEJ5	Q9lej5 glycine max
3	44	14.8	45	10 Q9LE39	Q9lej39 glycine max
4	44	14.8	56	16 Q9PDU4	Q9pdu4 xylella fas
5	43	14.5	53	5 Q9VQ65	Q9vq65 drosophila
6	43	14.5	56	5 Q91654	Q91654 ceratitis c
7	42	14.1	56	3 Q9UQY8	Q9uqy8 colletotric
8	41	13.8	44	2 Q9APQ5	Q9apq5 uncultured
9	41	13.8	48	2 Q9RID3	Q9rid3 yersinia pe
10	40	13.5	54	5 Q9V779	Q9v779 drosophila
11	39.5	13.3	56	10 Q9XFG5	Q9xfg5 vigna ungui
12	39	13.1	54	6 Q9WYS2	Q9mys2 oryctolagus
13	39	13.1	56	12 Q91G77	Q91g77 chilo iride
14	39	13.1	56	16 Q9K7C7	Q9k7c7 bacillus ha
15	39	13.1	57	5 Q16823	Q16823 drosophila
16	39	13.1	57	16 Q9PEJ8	Q9pej8 xylella fas

17	38	12.8	31	10 Q9LEJ3	Q9lej3 glycine max
18	38	12.8	37	5 Q9NLZ2	Q9nlz2 leishmania
19	38	12.8	37	6 Q18822	Q18822 sus scrofa
20	38	12.8	40	6 Q9TTD6	Q9ttid6 sus scrofa
21	38	12.8	48	2 Q02801	Q02801 streptomyce
22	38	12.8	50	5 Q27271	Q27271 plasmodium
23	38	12.8	53	16 Q922X1	Q922x1 rhizobium m
24	38	12.8	57	5 Q16824	Q16824 drosophila
25	37.5	12.6	53	2 Q93LH2	Q93lh2 sulfobact
26	37.5	12.6	55	4 Q9UM01	Q9umul homo sapien
27	37	12.5	42	2 Q9X8F1	Q9x8f1 streptomyce
28	37	12.5	46	4 Q15456	Q15456 homo sapien
29	37	12.5	50	5 Q26209	Q26209 plasmodium
30	37	12.5	50	5 Q26213	Q26213 plasmodium
31	37	12.5	55	5 Q9NL17	Q9nily7 leishmania
32	37	12.5	55	7 Q9MMF5	Q9mmf5 astatotilap
33	37	12.5	55	16 P76096	P76096 escherichia
34	36.5	12.3	39	16 Q9KDC1	Q9kdc1 bacillus ha
35	36.5	12.3	40	2 Q45678	Q45678 bacillus su
36	36.5	12.3	46	15 Q9NLZ9	Q9nlz9 mouse endog
37	36.5	12.3	55	10 Q9FFY2	Q9ffy2 arabidopsis
38	36	12.1	48	16 Q9ZL87	Q9z187 helicobacte
39	36	12.1	50	2 Q939R3	Q939r3 bacillus th
40	36	12.1	51	2 Q9R656	Q9r656 rhodobacter
41	36	12.1	53	12 Q56128	Q56128 porcine cir
42	36	12.1	53	12 Q91861	Q91861 porcine cir
43	36	12.1	57	5 Q9NM30	Q9nm30 leishmania
44	35.5	12.0	45	16 Q34405	Q34405 bacillus su
45	35.5	12.0	50	2 Q9KGZ4	Q9kgz4 clostridium

ALIGNMENTS

RESULT 1

Q9PA18	PRELIMINARY;	PRT;	54 AA.
ID	Q9PA18		
AC	Q9PA18;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	HYPOTHETICAL PROTEIN XF2529.		
GN	XF2529.		
OS	Xylella fastidiosa.		
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
OC	Xylella.		
OX	NCBI_TaxID=2371;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=9A5C;		
RX	MEDLINE=20365717; PubMed=10910347;		
RA	Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,		
RA	Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,		
RA	Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,		
RA	Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,		
RA	Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,		
RA	Coutinho L.B., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,		
RA	Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,		
RA	Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,		
RA	Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,		
RA	Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,		
RA	Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,		
RA	Lemos E.G.M., Lemus M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,		
RA	Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,		
RA	Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,		
RA	Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,		
RA	Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,		
RA	Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,		
RA	de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,		
RA	Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,		
RA	Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,		
RA	de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,		
RA	da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,		

```

RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL; AE004060; AAF85327.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 54 AA; 5931 MW; 25524B5C5018E13F CRC64;

Query Match 15.3%; Score 45.5; DB 16; Length 54;
Best Local Similarity 30.2%; Pred. No. 1.7e+02;
Matches 13; Conservative 7; Mismatches 22; Indels 1; Gaps 1;

QY 15 NAAADHAHAPE-DKYEAFVLHFSALHIIAGTPEVHAVK 56
| | | | | : : : : : | : : | |
DB 2 NNPSNEKHHTTPPLESLDECYSNFSLELIKALAAAYKEINAKKG 44

RESULT 2
Q9LEJ5 PRELIMINARY; PRT; 42 AA.
AC Q9LEJ5
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Stromvik M.V., Sundararaman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293440; CAB96762.1; -.
FT NON_TER 1
SQ SEQUENCE 42 AA; 4935 MW; 2327B258B8A2CFA5 CRC64;

Query Match 14.8%; Score 44; DB 10; Length 42;
Best Local Similarity 31.6%; Pred. No. 2e+02;
Matches 12; Conservative 6; Mismatches 12; Indels 8; Gaps 2;

QY 18 YNAADHAHAPE-DKYEAFVLHFSALHIIAGTPEVHAVK 54
| | | | | : : : : : | : : | |
DB 10 YEKVDHTAPEPTYKDLVVKLTKNV-----EAHLVE 40

RESULT 3
Q9LE39 PRELIMINARY; PRT; 45 AA.
AC Q9LE39
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG..
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Stromvik M.V., Sundararaman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293442; CAB96764.1; -.
DR EMBL; AJ293438; CAB96760.1; -.
FT NON_TER 1
SQ SEQUENCE 45 AA; 5236 MW; 2327E24CEA15458E CRC64;

Query Match 14.8%; Score 44; DB 10; Length 45;
Best Local Similarity 31.6%; Pred. No. 2.2e+02;
Matches 12; Conservative 6; Mismatches 12; Indels 8; Gaps 2;

QY 18 YNAADHAHAPE-DKYEAFVLHFSALHIIAGTPEVHAVK 54
| | | | | : : : : : | : : | |
DB 13 YEKVDHTAPEPTYKDLVVKLTKNV-----EAHLVE 43

RESULT 4
Q9PDU4 PRELIMINARY; PRT; 56 AA.
AC Q9PDU4
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN XF1285.
GN XF1285.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohnel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.B., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Martins E.M.F., Marinho C.L.,
RA Marques M.V., Martins A.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;

```


DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
FT NON_TER 1 56
ET NON_TER 56
SQ SEQUENCE 56 AA; 6688 MW; B2BA04C95819EAE CRC64;

Query Match 14.1%; Score 42; DB 3; Length 56;
Best Local Similarity 32.4%; Pred. No. 5e+02;
Matches 12; Conservative 6; Mismatches 17; Indels 2; Gaps 2;

Qy 4 POLVPK-LDEVYNAAYNAADHAAPEDKYEAFVLFHSE 39
|| : : : ||| | | | | : |
Db 16 PQIQNKIGISQLLGKAWNAESHEVRE-KYRALAKAYKE 51

RESULT 8

ID O9APQ5 PRELIMINARY; PRT; 44 AA.

AC O9APQ5;

DT 01-JUN-2001 (TremBLrel. 17, Created)

DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)

DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)

DE ELSS (FRAGMENT).

GN ELSS.

OS uncultured bacterium pCosHEL.

OC Bacteria; environmental samples.

OX NCBI_TaxID=143796;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20575196; PubMed=11133432;

RA Entcheva P., Liebl W., Johann A., Hartsch T., Streit W.R.;

RT "Direct cloning from enrichment cultures, a reliable strategy for

isolation of complete operons and genes from microbial consortia.";

RL Appl. Environ. Microbiol. 67:89-99(2001).

DR EMBL; AF250771; AAG60565.1; -

FT NON_TER 44 44

SQ SEQUENCE 44 AA; 5262 MW; 6C73B171FDE4F678 CRC64;

Query Match

Best Local Similarity 13.8%; Score 41; DB 2; Length 44;
Matches 9; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 11 DEVYNAAYNAADHAAPEDKYEAFV 34

||| : : : | : | : |

Db 18 DEYYPVAFDAIDRAQHKVLETFI 41

RESULT 9

ID Q9RID3

AC Q9RID3 PRELIMINARY; PRT; 48 AA.

DT 01-MAY-2000 (TremBLrel. 13, Created)

DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)

DE HYPOTHEICAL 5.4 KDA PROTEIN.

GN YMT1.42AC.

OS Yersinia pestis.

OC plasmid pMT1.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OX Yersinia.

OX NCBI_TaxID=632;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-CO-92 BIOVAR ORIENTALIS;

RA Baker S.G., Mungall K.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-CO-92 BIOVAR ORIENTALIS;

RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RP

RP

[3]

RN SEQUENCE FROM N.A.

RC STRAIN-CO-92 BIOVAR ORIENTALIS;

RA Karlyshev A.V., Wren B.W.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

KW EMBL; AL117211; CAB62369.1; -

KW Hypothetical protein; Plasmid.

SQ SEQUENCE 48 AA; 5417 MW; BFD2188AF24BED48 CRC64;

Query Match

Best Local Similarity 13.8%; Score 41; DB 2; Length 48;

Matches 11; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

Qy 10 LDEVYNAAYNAADHAAPEDKYEAFVLFHSEALHII 44

: | : | : | : | : | : |

Db 1 MSEFYSRAAAVADQAGNEDERAGPVFNLPPLIIV 35

RESULT 10

Q9V779

ID Q9V779 PRELIMINARY; PRT; 54 AA.

AC Q9V779;

DT 01-MAY-2000 (TremBLrel. 13, Created)

DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)

DE CG12853 PROTEIN.

GN CG12853.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang O., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,

RA Balleson K.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "Th2 genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003812; AAF58180.1; -.
DR FlyBase: FBgn040747; CG12853.
SQ SEQUENCE 54 AA; 5319 MW; DDDC6742335E8BAB CRC64;

Query Match 13.5%; Score 40; DB 5; Length 54;
Best Local Similarity 45.5%; Pred. No. 8.5e+02;
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 8 PKLDEVYNAAYNAADHAAPEDK 29
| | | | : | | | : | | |
Db 23 PTADPVTSESNAAPAAKEAK 44

RESULT 11

Q9XFG5 PRELIMINARY; PRT; 56 AA.
AC Q9XFG5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE GLUTELIN 2 (FRAGMENT).
OS Vigna unguiculata (Cowpea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
OX NCBI_TaxID=3917;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. I784E-124;
RA Machuka J.S.;
RT "Partial genomic nucleotide sequence of a glutelin 2 (gamma zein)
RT homolog from cowpea (Vigna unguiculata).";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF142332; AAD34914.1; -.
FT NON_TER 1
FT NON_TER 56
SQ SEQUENCE 56 AA; 6233 MW; D79613A9F8B84C08 CRC64;

Query Match 13.3%; Score 39.5; DB 10; Length 56;
Best Local Similarity 37.5%; Pred. No. 1e+03;
Matches 12; Conservative 4; Mismatches 11; Indels 5; Gaps 2;

Qy 26 PEDYKFAFLHFSALHTIAGTPE-VHAAKPG 56
| | : | | | : | | : | |
Db 26 PEPVHIPEPVHIPEVHI-----PEPVHIPEG 53

RESULT 12

Q9MYS2 PRELIMINARY; PRT; 54 AA.
AC Q9MYS2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEAT SHOCK PROTEIN 70 (FRAGMENT).
GN HSP70-1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND; TISSUE=BLOOD;
RA Rogel-Gallard C., Piumi F., Billault A., Bourgeaux N., Save J.C.,
RA Urien C., Salmon J., Chardon P.;
RT "Construction of a rabbit bacterial artificial chromosome (BAC)
RT library: application to the mapping of the major histocompatibility
RT complex to position 12q1.1";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ297379; CAB96070.1; -.
DR HSP: P04475; 1DKY.
FT NON_TER 1
FT NON_TER 54
SQ SEQUENCE 54 AA; 5748 MW; 44630BD1963DF3E1 CRC64;

Query Match 13.1%; Score 39; DB 6; Length 54;
Best Local Similarity 29.6%; Pred. No. 1.1e+03;
Matches 8; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

Qy 3 AFQLVKLDDEVYNAAYNAADHAAPEDK 29
| | | | : | | | : | | |
Db 23 APRGVQIEVTFDIDANGILHVTATDK 49

RESULT 13

Q91G77 PRELIMINARY; PRT; 56 AA.
AC Q91G77;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 018R.
OS Chilo iridescent virus (CIV) (Insect iridescent virus type 6).
OC Viruses; dsDNA viruses, no RNA stage; Iridoviridae; Iridovirus.
OX NCBI_TaxID=10488;
RN [1]
RP SEQUENCE FROM N.A.
RA Delius H., Darai G., Fluegel R.M.;
RT "DNA analysis of insect iridescent virus 6: evidence for circular
RT permutation and terminal redundancy.";
RL J. Virol. 49:609-614(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86174607; PubMed=3959991;
RA Lorbacher de Ruiz H., Gelderblom H., Hofmann W., Darai G.;
RT "Insect iridescent virus type 6 induced toxic degenerative hepatitis
RT in mice.";
RL Med. Microbiol. Immunol. 175:43-53(1986).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87321126; PubMed=2820141;
RA Schnitzler P., Soltan J.B., Fischer M., Reinsner H., Scholz J.,
RA Delius H., Darai G.;
RT "Molecular cloning and physical mapping of the genome of insect
RT iridescent virus type 6: further evidence for circular permutation of
RT the viral genome.";
RL Virology 160:66-74(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=89073752; PubMed=3201750;
RA Fischer M., Schnitzler P., Delius H., Darai G.;
RT "Identification and characterization of the repetitive DNA element in
RT the genome of insect iridescent virus type 6.";
RL Virology 167:485-496(1988).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92196996; PubMed=1549908;
RA Handermann M., Schnitzler P., Rosen-Wolff A., Raab K., Sonntag K.C.,
RA Darai G.;
RT "Identification and mapping of origins of DNA replication within the
RT DNA sequences of the genome of insect iridescent virus type 6";
RL Virus Genes 6:19-32(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=93118242; PubMed=1475907;
RA Sonntag K.C., Darai G.;
RT "Characterization of the third origin of DNA replication of the genome
RT of insect iridescent virus type 6.";
RL Virus Genes 6:333-342(1992).
RN [7]
RP SEQUENCE FROM N.A.

RX MEDLINE=93260401; PubMed=8492091;
RT Stohwasser R., Raab K., Schnitzler P., Janssen W., Darai G.;
RT "Identification of the gene encoding the major capsid protein of
RT insect iridescent virus type 6 by polymerase chain reaction.";
RL J. Gen. Virol. 74:873-879(1993).
RN [8]
RP SEQUENCE FROM N.A.
RX MEDLINE=94167241; PubMed=8121799;
RA Schnitzler P., Hug M., Handermann M., Janssen W., Koonin E.V.,
RA Delius H., Darai G.;
RT "Identification of genes encoding zinc finger proteins, non-histone
RT chromosomal HMG protein homologue, and a putative GTP phosphohydrolase
RT in the genome of Chilo iridescent virus.";
RL Nucleic Acids Res. 22:158-166(1994).
RN [9]
RP SEQUENCE FROM N.A.
RX MEDLINE=94353641; PubMed=8073636;
RA Sonntag K.C., Schnitzler P., Koonin E.V., Darai G.;
RT "Chilo iridescent virus encodes a putative helicase belonging to a
RT distinct family within the 'DEAD/H' superfamily: implications for the
RT evolution of large DNA viruses.";
RL Virus Genes 8:151-158(1994).
RN [10]
RP SEQUENCE FROM N.A.
RX MEDLINE=95213160; PubMed=7698884;
RA Sonntag K.C., Schnitzler P., Janssen W., Darai G.;
RT "Identification of the primary structure and the coding capacity of
RT the genome of insect iridescent virus type 6 between the genome
RT coordinates 0.310 and 0.347 (7990 bp).";
RL Intervirology 37:287-297(1994).
RN [11]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292906; PubMed=8021587;
RA Schnitzler P., Sonntag K.C., Muller M., Janssen W., Bugert J.J.,
RA Koonin E.V., Darai G.;
RT "Insect iridescent virus type 6 encodes a polypeptide related to the
RT largest subunit of eukaryotic RNA polymerase II.";
RL J. Gen. Virol. 75:1557-1567(1994).
RN [12]
RP SEQUENCE FROM N.A.
RX MEDLINE=98141693; PubMed=9482589;
RA Bahr U., Tidona C.A., Darai G.;
RT "The DNA sequence of Chilo iridescent virus between the genome
RT coordinates 0.101 and 0.391; similarities in coding strategy between
RT insect and vertebrate iridoviruses.";
RL Virus Genes 15:235-245(1997).
RN [13]
RP SEQUENCE FROM N.A.
RX MEDLINE=99125223; PubMed=9926400;
RA Muller K., Tidona C.A., Bahr U., Darai G.;
RT "Identification of a thymidylate synthase gene within the genome of
RT Chilo iridescent virus.";
RL Virus Genes 17:243-258(1998).
RN [14]
RP SEQUENCE FROM N.A.
RX MEDLINE=99383793; PubMed=10456793;
RA Muller K., Tidona C.A., Darai G.;
RT "Identification of a gene cluster within the genome of Chilo
RT iridescent virus encoding enzymes involved in viral DNA replication
RT and processing.";
RL Virus Genes 18:243-264(1999).
RN [15]
RP SEQUENCE FROM N.A.
RX MEDLINE=21342589; PubMed=11448171;
RA Jakob N.J., Muller K., Bahr U., Darai G.;
RT "Analysis of the First Complete DNA Sequence of an Invertebrate
RT Iridovirus: Coding Strategy of the Genome of Chilo Iridescent Virus.";
RL Virology 286:182-196(2001).
RN [16]
RP SEQUENCE FROM N.A.
RA Jakob N.J., Mueller K., Bahr U., Darai G.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF303741; AAK81955.1; -

SQ SEQUENCE 56 AA; 7186 MW; DE0A2E74201055F9 CRC64;
Query Match 13.1%; Score 39; DB 12; Length 56;
Best Local Similarity 25.8%; Pred. No. 1.2e+03;
Matches 8; Conservative 8; Mismatches 9; Indels 6; Gaps 2;
QY 19 NAADHAAPEDKYEAF-----VLH-FSEALHI 43
DB 16 NFNHNPPQNVLYSHYQLIIHQYTDYCHV 46
RESULT 14
QYK7C7 ID Q9K7C7 PRELIMINARY; PRT; 56 AA.
AC Q9K7C7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DE 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE BH3437 PROTEIN.
GN BH3437.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AF001518; BAB07156.1; -
KW Complete proteome.
SQ SEQUENCE 56 AA; 6569 MW; 2CFE84E462446B3C CRC64;
Query Match 13.1%; Score 39; DB 16; Length 56;
Best Local Similarity 37.5%; Pred. No. 1.2e+03;
Matches 9; Conservative 2; Mismatches 13; Indels 0; Gaps 0;
QY 10 LDEVYNAAYNAADHAAPEDKYEAF 33
DB 17 LDEVYNGRRPAPHKMTKIQLY 40
RESULT 15
O16823 ID O16823 PRELIMINARY; PRT; 57 AA.
AC O16823;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ANDROPIN.
GN ANP OR CG1361.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B316;
RX MEDLINE=97476321; PubMed=9335607;
RA Clark A.G., Wang L.;
RT "Molecular population genetics of Drosophila immune system genes.";
RL Genetics 147:713-724(1997).
DR EMBL; AF018988; AAB82485.1; -
DR FlyBase; FBgn000094; Anp.
SQ SEQUENCE 57 AA; 6185 MW; F2C57A0859B839B5 CRC64;

Query Match. 13.1%; Score 39; DB 5; Length 57;
Best Local Similarity 45.8%; Pred. No. 1.2e+03;
Matches 11; Conservative 3; Mismatches 6; Indels 4; Gaps 1;

Qy 10 LDEVNAAAYNAAD---HAAPEDK 29
||:| || :||| | | :|
Db 29 LDKVENAIHNAAQVGVGFAKPFK 52

Search completed: July 11, 2002, 10:59:22
Job time: 210 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:52:37 ; Search time 12.96 seconds
(without alignments)
107.427 Million cell updates/sec

Title: US-09-696-169A-19
Perfect score: 297
Sequence: 1 SKAPQLVPKLDVYNAAYNA.....SEALHIIAGTPEVHAVKGA 57

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 166652

Minimum DB seq length: 0
Maximum DB seq length: 57

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result		Query		SUMMARIES		Description	
No.	Score	Match	Length	DB	ID		
1	65	21.9	20	1	US-08-440-861-12	Sequence 12, Appl	Sequence 12, Appl
2	62	20.9	20	1	US-08-440-861-11	Sequence 11, Appl	Sequence 11, Appl
3	56	18.9	20	1	US-08-440-861-13	Sequence 13, Appl	Sequence 13, Appl
4	54	18.2	20	1	US-08-440-861-14	Sequence 14, Appl	Sequence 14, Appl
5	48	16.2	13	1	US-08-433-854-28	Sequence 28, Appl	Sequence 28, Appl
6	48	16.2	13	1	US-08-174-745A-28	Sequence 28, Appl	Sequence 28, Appl
7	48	16.2	13	2	US-08-195-947-28	Sequence 28, Appl	Sequence 28, Appl
8	48	16.2	13	2	US-08-433-885-28	Sequence 28, Appl	Sequence 28, Appl
9	48	16.2	13	2	US-08-433-908B-28	Sequence 28, Appl	Sequence 28, Appl
10	48	16.2	13	4	US-08-410-614-28	Sequence 28, Appl	Sequence 28, Appl
11	48	16.2	20	1	US-08-440-861-15	Sequence 15, Appl	Sequence 15, Appl
12	44	14.8	20	1	US-08-440-861-10	Sequence 10, Appl	Sequence 10, Appl
13	43.5	14.6	36	1	US-08-214-770-8	Sequence 8, Appl	Sequence 8, Appl
14	43.5	14.6	36	5	PCT-US95-02885-8	Sequence 8, Appl	Sequence 8, Appl
15	41	13.8	20	1	US-08-440-861-8	Sequence 8, Appl	Sequence 8, Appl
16	40	13.5	50	4	US-08-963-851-9	Sequence 9, Appl	Sequence 9, Appl
17	40	13.5	53	4	US-08-900-574-4	Sequence 4, Appl	Sequence 4, Appl
18	39.5	13.3	43	1	US-08-178-477B-21	Sequence 21, Appl	Sequence 21, Appl
19	39	13.1	20	1	US-08-440-861-7	Sequence 7, Appl	Sequence 7, Appl
20	38.5	13.0	35	2	US-08-749-526-6	Sequence 6, Appl	Sequence 6, Appl
21	38	12.8	20	3	US-08-817-926-3	Sequence 3, Appl	Sequence 3, Appl
22	38	12.8	32	4	US-09-082-279B-1226	Sequence 1226, Ap	Sequence 1226, Ap
23	38	12.8	32	4	US-09-315-304B-1226	Sequence 1226, Ap	Sequence 1226, Ap
24	37	12.5	35	4	US-09-082-279B-784	Sequence 784, App	Sequence 784, App
25	37	12.5	35	4	US-09-315-304B-784	Sequence 784, App	Sequence 784, App
26	37	12.5	45	4	US-08-963-851-10	Sequence 10, Appl	Sequence 10, Appl
27	37	12.5	47	4	US-08-936-165A-365	Sequence 365, App	Sequence 365, App

28	37	12.5	48	1	US-08-446-692-37	Sequence 37, Appl
29	37	12.5	48	2	US-08-488-351A-37	Sequence 37, Appl
30	36	12.1	12	1	US-08-433-854-23	Sequence 23, Appl
31	36	12.1	12	1	US-08-433-854-25	Sequence 25, Appl
32	36	12.1	12	1	US-08-174-745A-23	Sequence 23, Appl
33	36	12.1	12	1	US-08-174-745A-25	Sequence 25, Appl
34	36	12.1	12	2	US-08-195-947-23	Sequence 23, Appl
35	36	12.1	12	2	US-08-195-947-25	Sequence 25, Appl
36	36	12.1	12	2	US-08-433-885-23	Sequence 23, Appl
37	36	12.1	12	2	US-08-433-885-25	Sequence 25, Appl
38	36	12.1	12	2	US-08-433-908B-23	Sequence 23, Appl
39	36	12.1	12	2	US-08-433-908B-25	Sequence 25, Appl
40	36	12.1	12	4	US-08-410-614-23	Sequence 23, Appl
41	36	12.1	12	4	US-08-410-614-25	Sequence 25, Appl
42	36	12.1	21	4	US-08-423-646A-66	Sequence 66, Appl
43	36	12.1	30	4	US-09-099-307-3	Sequence 3, Appl
44	36	12.1	38	1	US-08-176-500-45	Sequence 45, Appl
45	36	12.1	38	1	US-08-471-052A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-440-861-12
: Sequence 12, Application US/08440861
: Patent No. 5710126
: GENERAL INFORMATION:
: APPLICANT: Griffith, Irwin J.
: APPLICANT: Kuo, Mei-Chang
: APPLICANT: Luqman, Mohammad
: TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
: TITLE OF INVENTION: ALLERGEN
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440,861
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/106,016
: FILING DATE: 31-AUG-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Amy E. Mandragouras
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 20 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: internal
: US-08-440-861-12

Query Match 21.9%; Score 65; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.023;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;


```

QY 14 YNAAYNAADHAAPEAKYEF 33
; Sequence 11, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Lugman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-11

Query Match 20.9%; Score 62; DB 1; Length 20;
Best Local Similarity 57.9%; Pred. No. 0.057;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 4 POLVPKLDEVYNAAYNAAD 22
; Sequence 13, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Lugman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-11

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-13

Query Match 18.9%; Score 56; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. No. 0.36;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 24 AAPEDKYEAFVLHFSEALHI 43
; Sequence 14, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Lugman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-13

Db 1 ATPEAKYDAFVTALTEALRV 20
; Sequence 14, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Lugman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-13

```

```

; APPLICATION NUMBER: US 08/106,016
;
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-0400cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-14

```

Query Match 18.2%; Score 54; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. NO. 0.66;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY	34	VLHFSEALHIIAGTPEVHAV	53
		: :	
Db	1	VTALTEALRVIAGALEVHAV	20

```

RESULT      5
US-08-433-854-28,
; Sequence 28, Application US/08433854
; Patent No. 5721119
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjotglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:

```

```
,  
ZLIB:      19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE:   Floppy disk  
COMPUTER:    IBM PC compatible  
OPERATING SYSTEM:  PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/433,854
```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid

```

; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-854-28

Query Match 16.2%; Score 48; DB 1; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 0;

QY 44 IAGTPEVHAVKP 55
    ||| |||||
Db 1 IAGALEVHAVKP 12

RESULT 6
US-08-174-745A-28
; Sequence 28, Application US/08174745A
; Patent No. 5736362
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Koox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphiglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5736362rls
; STREET: One Liberty Place -46th Floor

```

Query Match	16.2%	Score 48;	DB 1;	Length 13;
Best Local Similarity	83.3%	Pred. No. 2.4;		
Matches 10; Conservative		0; Mismatches	2; Indels	
QY	44	IAGTPEVHVAYKP	55	
Db	1	IAGALEVHVAYKP	12	

```
RESULT 7
US-08-195-947-28
; Sequence 28, Application US/08195947
; Patent No. 5840316
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjiloglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,947
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-195-947-28

Query Match 16.2%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55
Db 1 IAGALEVHAVKP 12

RESULT 8
US-08-433-885-28
; Sequence 28, Application US/08433885
; Patent No. 5869333
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjiloglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
```

```
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,885
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-885-28

Query Match 16.2%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55
Db 1 IAGALEVHAVKP 12

RESULT 9
US-08-433-908B-28
; Sequence 28, Application US/08433908B
; Patent No. 5965455
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjiloglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,908B
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 436
```

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-908B-28

Query Match 16.2%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55
||| |||||
DB 1 IAGALEVHAVKP 12

RESULT 10
US-08-410-614-28
; Sequence 28, Application US/08410614
; Patent No. 6277383
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjiloglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Genk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,614
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/195,947
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

MOLECULE TYPE: peptide
US-08-410-614-28

Query Match 16.2%; Score 48; DB 4; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.4;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55
||| |||||
DB 1 IAGALEVHAVKP 12

RESULT 11
US-08-440-861-15
; Sequence 15, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragoras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-15

Query Match 16.2%; Score 48; DB 1; Length 20;
Best Local Similarity 83.3%; Pred. No. 4.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55
||| |||||
DB 1 IAGALEVHAVKP 12

RESULT 12
US-08-440-861-10
; Sequence 10, Application US/08440861
; Patent No. 5710126

```

: GENERAL INFORMATION:
: APPLICANT: Griffith, Irwin J.
: APPLICANT: Kuo, Mei-Chang
: APPLICANT: Luqman, Mohammad
: TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
: TITLE OF INVENTION: ALLERGEN
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 State Street, suite 510
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/440,861
: FILING DATE: 15-MAY-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/106,016
: FILING DATE: 31-AUG-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Amy E. Mandragouras
: REGISTRATION NUMBER: 36,207
: REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-7400
: TELEFAX: (617) 227-5941
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 20 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: FRAGMENT TYPE: Internal
: US-08-440-861-10

```

Query Match 14.8%; Score 44; DB 1; Length 20;
Best Local Similarity 72.7%; Pred. NO. 14;
Matches 8; Conservative 2; Mismatches 1; Indels

Qy 1 SKAPQLVPKLD 11
:| | | | :| | | |
Db 8 AKAPGLIPKLD 18

```

RESULT 13
US-08-214-770-8
; Sequence 8, Application US/08214770
; Patent No. 5523209
; GENERAL INFORMATION:
; APPLICANT: Ginsberg, Mark H.
; APPLICANT: O'Toole, Tim
; TITLE OF INVENTION: METHODS FOR IDENTIFYING
; TITLE OF INVENTION: INHIBITORS OF INTEGRIN
; TITLE OF INVENTION: ACTIVATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; CITY: Boston
; STREET: 225 Franklin Street
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX

```

```

; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: WordPerfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,770
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06410/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-214-770-8

```

Query Match 14.6%; Score 43.5; DB 1; Length 36;
Best Local Similarity 52.9%; Pred. NO. 34;
Matches 9; Conservative 3; Mismatches 4; Indels

QY 14 YNAYNAAD-HAAPEDK 29
| : | : | : | : | : | :
Db 13 YDATYHKAIEIHAQPSDK 29

RESULT 14
PCT-US95-02885-8
; Sequence 8, Application PC/TUS9502885
; GENERAL INFORMATION:
; APPLICANT: Ginsberg, Mark H.
; APPLICANT: O'Toole, Timothy
; TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS
; TITLE OF INVENTION: OF INTEGRIN ACTIVATION
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: MS-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02885
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/214,770
; FILING DATE: March 14, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06410/002001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 36
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US95-02885-8

Query Match 14.6% Score 43.5; DB 5; Length 36;
Best Local Similarity 52.9%; Pred. No. 34;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 14 YNAAYNAAAD-HAAPEDK 29
| | | | | | | | | |
Db 13 YDATYHKAETHAQPSDK 29

RESULT 15
US-08-440-861-8
; Sequence 8, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-8

Query Match 13.8% Score 41; DB 1; Length 20;
Best Local Similarity 44.4%; Pred. No. 35;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 23 HAAPEDKYEAFVLHFSEA 40
| | | | | | | | | |
Db 1 NAPPADFKIFEAFSES 18

Search completed: July 11, 2002, 10:56:11
Job time: 214 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	40	14.4	36	1	YYPIS	peptide YV - short
2	40	14.4	49	2	S70651	leukotriene-A4 hyd
3	39	14.1	43	2	S52358	hypothetical prote
4	39	14.1	52	1	R3KM72	ribosomal protein
5	38.5	13.9	49	1	GECT	osteocalcin - cat
6	38	13.7	41	2	S50084	single stranded DN
7	38	13.7	53	2	G95302	hypothetical prote
8	37	13.4	34	2	B41397	hypothetical prote
9	37	13.4	37	2	A26781	peptide YG - Ameri
10	37	13.4	42	2	T36238	hypothetical prote
11	37	13.4	46	2	S38225	protein-serine/thr
12	36.5	13.2	27	2	A33210	protein-disulfide-
13	36.5	13.2	39	2	D83811	hypothetical prote
14	36.5	13.2	52	2	AC2208	hypothetical prote
15	36	13.0	48	2	E71901	hypothetical prote
16	35.5	12.8	45	2	A69871	conserved hypothet
17	35	12.6	36	2	D69141	hypothetical prote
18	35	12.6	42	2	A95019	hypothetical prote
19	35	12.6	43	2	A42852	meia protein - sal
20	35	12.6	46	2	A83629	hypothetical prote
21	35	12.6	46	2	T50586	hypothetical prote
22	34.5	12.5	47	2	S24142	cholecystokinin/ga
23	34.5	12.5	48	2	S02208	osteocalcin - emu
24	34.5	12.5	49	1	GEMK1	osteocalcin - crab
25	34.5	12.5	53	1	A61280	osteocalcin - rabb
26	34.5	12.5	53	2	B87570	hypothetical prote
27	34	12.3	39	2	A82359	hypothetical prote
28	34	12.3	42	2	I37543	MHC HLA-DR-beta-1
29	34	12.3	48	1	R5VF32	ribosomal protein

RESULT 2
S70651
leukotriene-A4 hydrolase (EC 3.3.2.6) short isoform - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 07-May-1999
C:Accession: S70651
R:Jendraschak, E.; Kaminski, W.E.; Kiefl, R.; von Schacky, C.
Biochem. J. 314, 733-737, 1996
A:Title: The human leukotriene A(4) hydrolase gene is expressed in two alternatively
A:Reference number: S70627; MUID:96177841
A:Accession: S70651
A:Molecule type: DNA; mRNA
A:Residues: 1-49 <JEN>

Matches 9; Conservative 3; Mismatches 5; Indels 3; Gaps 1;

Qy 21 APEDKYEAFVLH---FSRAL 37
||||| |

Db 2 APDEEDHVLVLHKGNFDEAL 21

RESULT 13

D83811

hypothetical protein BH1292 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83811

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83811

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-39 <STO>

A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA05011.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH1292

Query Match 13.2%; Score 36.5; DB 2; Length 39;

Best Local Similarity 32.3%; Pred. No. 6.8e+02;

Matches 10; Conservative 7; Mismatches 9; Indels 5; Gaps 2;

Qy 5 KLDEVYNAAY--NAA---DHAAPEDKYEAFV 30

||||| : | | | | | | | |

Db 2 KANDLYKNEFTISNGAEWNDFAQEAQYQSFI 32

RESULT 14

AC2208

hypothetical protein asr3218 [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AC2208

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2208

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-52 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA074917.1; PID:g17132313; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: asr3218

Query Match 13.2%; Score 36.5; DB 2; Length 52;

Best Local Similarity 32.0%; Pred. No. 9.4e+02;

Matches 8; Conservative 5; Mismatches 7; Indels 5; Gaps 1;

Qy 27 EAFVLH-----FSEALHIIAGTPEV 46

||| : | | | | | | | |

Db 19 EYLLHNGRKNLPEPIHLVRKTPM 43

RESULT 15

E71901

hypothetical protein jhp0693 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: E71901

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D. ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557

A:Accession: E71901

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-48 <ARN>

A:Cross-references: GB:AE001500; GB:AE001439; NID:g4155238; PIDN:AAD06268.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp0693

Query Match 13.0%; Score 36; DB 2; Length 48;

Best Local Similarity 34.5%; Pred. No. 9.9e+02;

Matches 10; Conservative 3; Mismatches 12; Indels 4; Gaps 1;

Qy 7 DEVYNAAYNAADHAAPEDKYEAFVLHFSE 35

||||| : | | | | | | | |

Db 23 DEEYNDYKNVYD----DDDYEDYNSDYEE 47

Search completed: July 11, 2002, 11:00:38

Job time: 130 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:00:43 ; Search time 10.16 Seconds
(without alignments)
201.982 Million cell updates/sec

Title: US-09-696-169A-20

Perfect score: 277

Sequence: 1 QLVPKLDEVYNAAYNAADHA.....SEALHIIAGTPEVHAVKPGA 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 3990

Minimum DB seq length: 0

Maximum DB seq length: 53

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	14.4	36	1 PYI_MYOSC	P09641 myoxocephal
2	40	14.4	36	1 PYI_RAJRH	P29206 raja rhina
3	39	14.1	36	1 PYI_ORENI	P81028 oreochromis
4	38.5	13.9	49	1 OSTC_FELCA	P02821 felis silve
5	37.5	13.5	49	1 OSTC_CANFA	P81455 canis famli
6	34.5	12.5	48	1 OSTC_DRONO	P15504 dromailus no
7	34.5	12.5	49	1 OSTC_MACFA	P02819 macaca fasc
8	34.5	12.5	49	1 OSTC_RABIT	P39056 oryctolagus
9	34.5	12.5	49	1 OSTC_XENLA	P40147 xenopus lae
10	34	12.3	46	1 PSBK_SYNEL	Q9fik9 synechococ
11	34	12.3	48	1 RK32_VICFA	P15820 vicia faba
12	34	12.3	51	1 INS_BALBO	P01314 balaenopter
13	34	12.3	51	1 INS_BALPH	P01312 balaenopter
14	34	12.3	51	1 INS_CAMDR	P01320 camelus dro
15	34	12.3	51	1 INS_CAPHI	P01319 capra hircu
16	34	12.3	51	1 INS_DIDMA	P18109 didelphis m
17	34	12.3	51	1 INS_ELEMA	P01316 elephas max
18	34	12.3	51	1 INS_FELCA	P06306 felis silve
19	33.5	12.1	51	1 YNFN_ECOLI	P76157 escherichia
20	33	11.9	51	1 INS_ANGRO	P42633 anguilla ro
21	32.5	11.7	39	1 FUC3_RAT	P80349 rattus norv
22	32	11.6	50	1 INS_KATPE	P01340 katsuwonus
23	32	11.6	51	1 INS_ACOCA	P01324 acomyus cahi
24	32	11.6	51	1 INS_ANGAN	P07454 anser anser
25	32	11.6	51	1 INS_HYSCR	P01328 hystrix cri
26	32	11.6	51	1 INS_ORNAN	Q9tcy7 ornithorhyn
27	32	11.6	51	1 INS_TRASC	P31887 trachemys s
28	32	11.6	52	1 INS_ACTIGU	P81423 acipenser g
29	32	11.6	52	1 INS_LEPSP	P09476 lepisosteus
30	32	11.6	53	1 TWHH_CARAU	P79696 carassius a
31	32	11.6	53	1 TWHH_DANKE	P79715 danio kerri
32	32	11.6	53	1 TWHH_DEVPA	O13253 devario pat
33	32	11.6	53	1 TWHH_PUNTE	P79855 puntius tet

ALIGNMENTS

RESULT 1

PYI_MYOSC

ID PYI_MYOSC STANDARD; PRT; 36 AA.

AC P09641;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Peptide YY-like (PYI).

OS Myoxocephalus scorpius (Shorthorn sculpin) (Daddy sculpin).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Scorpaeniformes;

OC Cottidae; Cottidae; Myoxocephalus.

OX NCBI_TaxID=8097;

RN [1]

RP SEQUENCE.

RC TISSUE=Pancreas;

RX MEDLINE=87190954; PubMed=2883025;

RA Cutfield S.M., Carne A., Cutfield J.F.;

RT "The amino-acid sequences of sculpin islet somatostatin-28 and

peptide YY.";

RL FEBS Lett. 214:57-61(1987).

RN [2]

RP SEQUENCE.

RX MEDLINE=87176585; PubMed=3562898;

RA Conlon J.M., Schmidt W.E., Gallwitz B., Falkmer S., Thim L.;

RT "Characterization of an amidated form of pancreatic polypeptide from

the daddy sculpin (Cottus scorpius).";

RL Regul. Pept. 16:261-268(1986).

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: BELONGS TO THE NPY / PPY / PYI FAMILY.

DR PIR; A60309; YVEIS.

DR HSSP; P01303; INON.

DR InterPro; IPR001955; Pancreatic_hormn.

DR Pfam; PF00159; hormone3; 1.

DR PRINTS; PR00278; PANCHORMONE.

DR ProDom; PD001267; Pancreatic_hormn; 1.

DR SMART; SM00309; PAH; 1.

DR PROSITE; PS00265; PANCREATIC_HORMONE_1; 1.

DR PROSITE; PS0276; PANCREATIC_HORMONE_2; 1.

KW Hormone; Amidation.

FT MOD_RES 36 36

SQ SEQUENCE 36 AA; 4170 MW; 16F331B73643D7AA CRC64;

AMIDATION.

Query Match 14.4%; Score 40; DB 1; Length 36;

Best Local Similarity 52.9%; Pred. No. 79;

Matches 9; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 19 HAAPED--KYAEAFVLHF 33

:|:|:| | | | |

Db 11 NASPEDWAKYHAAVRHY 27

RESULT 2

PYI_RAJRH

```

ID PYV_RAJRH STANDARD; PRT; 36 AA.
AC P29206;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptide YY-like (PYV).
OS Raja rhina (Skate).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Squala; Hypnosquales; Pristiogaster; Batoidae;
OC Rajiformes; Rajidae; Raja.
OX NCBI_TaxID=30478;
RN [1]
RP SEQUENCE.
RX MEDLINE=91296574; PubMed=2067973;
RA Conlon J.M., Bjerning C., Moon T.W., Youson J.H., Thim L.;
RT "Neuropeptide Y-related peptides from the pancreas of a teleostean
RL (eel), holostean (bowfin) and elasmobranch (skate). fish.";
RL Peptides 12:221-226(1991).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE NPV / PPV / PYV FAMILY.
DR HSSP; P01303; IRON.
DR InterPro; IPR001955; Pancreatic_hormn.
DR Pfam; PF00159; hormone3; 1.
DR PRINTS; PR00278; PANCHORMONE.
DR ProDom; PD001267; Pancreatic_hormn; 1.
DR SMART; SM00309; PAH; 1.
DR PROSITE; PS00265; PANCREATIC_HORMONE_1; 1.
DR PROSITE; PS00276; PANCREATIC_HORMONE_2; 1.
KW Hormone; Amidation.
FT MOD_RES 36
SQ SEQUENCE 36 AA; 4251 MW; 07A7D9DC196660B6 CRC64;

Query Match 14.4%; Score 40; DB 1; Length 36;
Best Local Similarity 42.9%; Pred. No. 79;
Matches 9; Conservative 4; Mismatches 6; Indels 2; Gaps 1;

QY 15 NAADHAPEP--KYFAFVLHF 33
| | |||| | : : |
DB 7 NPGDDAAPEELAKYYSALRHY 27

RESULT 3
PYV_ORENI
ID PYV_ORENI STANDARD; PRT; 36 AA.
AC P81028;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peptide YY-like (PYV).
OS Oreochromis niloticus (Nile tilapia) (Tilapia nilotica).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Labroidel;
OC Cichlidae; Oreochromis.
OX NCBI_TaxID=8128;
RN [1]
RP SEQUENCE.
RX MEDLINE=95384941; PubMed=7656183;
RA Nguyen T.M., Wright J.R. Jr., Nielsen P.F., Conlon J.M.;
RT "Characterization of the pancreatic hormones from the Brockmann body
of the tilapia: implications for islet xenograft studies.";
RL Comp. Biochem. Physiol. 111C:33-44(1995).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE NPV / PPV / PYV FAMILY.
DR HSSP; P01303; IRON.
DR InterPro; IPR001955; Pancreatic_hormn.
DR Pfam; PF00159; hormone3; 1.
DR PRINTS; PR00278; PANCHORMONE.
DR ProDom; PD001267; Pancreatic_hormn; 1.
DR SMART; SM00309; PAH; 1.
DR PROSITE; PS00265; PANCREATIC_HORMONE_1; 1.
DR PROSITE; PS00276; PANCREATIC_HORMONE_2; 1.

```

```

KW Hormone; Amidation. 36
FT MOD_RES 36
SQ SEQUENCE 36 AA; 4201 MW; 0246CFBC6243D7AA CRC64;

Query Match 14.1%; Score 39; DB 1; Length 36;
Best Local Similarity 56.2%; Pred. No. 11e+02;
Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 20 AAPED--KYFAFVLHF 33
| : ||| | | : |
DB 12 ASPEDWAKYHAAVRHY 27

RESULT 4
OSTC_FELCA
ID OSTC_FELCA STANDARD; PRT; 49 AA.
AC P02821;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
DE protein) (BGP).
GN BGLAP.
OS Felis silvestris catus (Cat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX NCBI_TaxID=9685;
RN [1]
RP SEQUENCE.
RX MEDLINE=85054706; PubMed=6334077;
RA Shimomura H., Kanai Y., Sanada K.;
RT "Primary structure of cat osteocalcin.";
RL J. Biochem. 96:405-411(1984).
CC -1- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC STRONGLY TO APATITE AND CALCIUM.
CC -1- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -1- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR PIR; A03304; GECT.
DR InterPro; IPR002384; GLA_bone.
DR InterPro; IPR000294; Vitk_dep_GLA.
DR Pfam; PF00594; gla; 1.
DR PRINTS; PR00002; GLABONE.
DR SMART; SM00069; GLA; 1.
DR PROSITE; PS00011; GLU_CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES 9
FT MOD_RES 9 HYDROXYLATION.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29
FT SEQUENCE 49 AA; 5495 MW; 93D2131FA9F656D3 CRC64;

Query Match 13.9%; Score 38.5; DB 1; Length 49;
Best Local Similarity 34.5%; Pred. No. 1.7e+02;
Matches 10; Conservative 3; Mismatches 9; Indels 7; Gaps 1;

QY 1 QLVPKLDEVNAYNAADHAAPEDKYEAF 29
: | | | | : | | |
DB 24 ELNPDCDEL-----ADHIGFQDAYRRF 45

RESULT 5
OSTC_CANFA
ID OSTC_CANFA STANDARD; PRT; 49 AA.
AC P81455;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)

```



```
Best Local Similarity 31.0%; Pred. No. 5.6e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

QY 1 QLVPKLDEVYNAAYNAADHAAPDKYEAF 29
   :| | ||:
Db 24 ELNPDCEL-----ADHIGFQAYRRF 45

RESULT 8
OSTC_RABIT
ID OSTC_RABIT STANDARD; PRT; 49 AA.
AC P39056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein)
DE (Bone Gla-protein) (BGP).
GN BGLAP.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=92175242; PubMed=1794506;
RT "Primary aminoacid sequence of rabbit osteocalcin.";
RL Biochem. Soc. Trans. 19:373S-373S(1991).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC -!- STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR PIR: A61280; A61280.
DR InterPro: IPR002384; GLA_bone.
DR InterPro: IPR000294; VitK_dep_GLA.
DR Pfam: PF00594; gla; 1.
DR PRINTS: PR00002; GLABONE.
DR SMART: SM00069; GLA; 1.
DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Hydroxylation;
KW Bone.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 9 9 HYDROXYLATION.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29 BY SIMILARITY.
SQ SEQUENCE 49 AA; 5431 MW; 7B218871F0312253 CRC64;

Query Match 12.5%; Score 34.5; DB 1; Length 49;
Best Local Similarity 31.0%; Pred. No. 5.6e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

QY 1 QLVPKLDEVYNAAYNAADHAAPDKYEAF 29
   :| | ||:
Db 24 ELNPDCEL-----ADQVGLQAYQRF 45

RESULT 9
OSTC_XENLA
ID OSTC_XENLA STANDARD; PRT; 49 AA.
AC P40147;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Osteocalcin (Gamma-carboxyglutamic acid-containing protein) (Bone Gla-
DE protein) (BGP).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE.
RX MEDLINE=96139691; PubMed=8567186;
RA Cancela M., Williamson M.K., Price P.A.;
RT "Amino-acid sequence of bone Gla protein from the African clawed toad
RT Xenopus laevis and the fish Sparus aurata.";
RL Int. J. Pept. Protein Res. 46:419-423(1995).
CC -!- FUNCTION: CONSTITUTES 1-2% OF THE TOTAL BONE PROTEIN. IT BINDS
CC -!- STRONGLY TO APATITE AND CALCIUM.
CC -!- PTM: GAMMA-CARBOXYGLUTAMIC ACID RESIDUES ARE FORMED BY VITAMIN K
CC DEPENDENT CARBOXYLATION. THESE RESIDUES ARE ESSENTIAL FOR THE
CC BINDING OF CALCIUM.
CC -!- SIMILARITY: BELONGS TO THE OSTEOCALCIN / MATRIX GLA-PROTEIN
CC FAMILY.
DR InterPro: IPR002384; GLA_bone.
DR InterPro: IPR000294; VitK_dep_GLA.
DR Pfam: PF00594; gla; 1.
DR PRINTS: PR00002; GLABONE.
DR SMART: SM00069; GLA; 1.
DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
KW Calcium-binding; Gamma-carboxyglutamic acid; Vitamin K; Bone.
FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
FT MOD_RES 24 24 GAMMA-CARBOXYGLUTAMIC ACID.
FT DISULFID 23 29 BY SIMILARITY.
SQ SEQUENCE 49 AA; 5360 MW; 7A9A8F63A12E6047 CRC64;

Query Match 12.5%; Score 34.5; DB 1; Length 49;
Best Local Similarity 31.0%; Pred. No. 5.6e+02;
Matches 9; Conservative 4; Mismatches 9; Indels 7; Gaps 1;

QY 1 QLVPKLDEVYNAAYNAADHAAPDKYEAF 29
   :| | ||:
Db 24 ELNPDCEL-----ADHIGFQAYRRF 45

RESULT 10
PSBK_SYNEL
ID PSBK_SYNEL STANDARD; PRT; 46 AA.
AC Q9FIK9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Photosystem II reaction center protein K precursor (PSII-K).
GN PSBK.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RA Katoh H., Ikeuchi M.;
RT "Cloning and disruption of the psbk gene from thermophilic
RT Thermosynechococcus elongatus.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS A COMPONENT OF THE REACTION CENTER
CC OF PHOTOSYSTEM II.
CC -!- SIMILARITY: BELONGS TO THE PSBK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AB052850; BAB20628.1;
CC InterPro: IPR003687; Psbk.
CC Pfam: PF02533; Psbk; 1.
```

```
KW Photosystem II.
FT PROPEP 1 9 BY SIMILARITY.
FT CHAIN 10 46 PHOTOSYSTEM II REACTION CENTER PROTEIN K.
SQ SEQUENCE 46 AA; 5026 MW; 45C5197F4B50E398 CRC64;

Query Match 12.3%; Score 34; DB 1; Length 46;
Best Local Similarity 40.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 35 EALHIIAGTPEVHAV 49
DB 3 DALVLVAKLPEAYAI 17

RESULT 11
RK32_VICFA STANDARD; PRT; 48 AA.
AC P15820;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Chloroplast 50S ribosomal protein L32.
GN RPL32.
OS Vicia faba (Broad bean).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Vicia.
OX NCBI_TaxID=3906;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206803; PubMed=2320425;
RA Hardenberger F., Pillay D.F.N., Steinmetz A.;
RT "Sequence of the trnH gene and the inverted repeat structure deletion
RL site of the broad bean chloroplast genome."
RL Nucleic Acids Res. 18:1297-1297(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89208884; PubMed=3242868;
RA Hardenberger F., Weil J.H., Steinmetz A.;
RT "Organization and nucleotide sequence of the broad bean chloroplast
RT genes trnL-UAG, ndhF and two unidentified open reading frames."
RL Curr. Genet. 14:609-615(1988).
CC -1- SIMILARITY: BELONGS TO THE L32P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X51471; CAA35833.1; -.
DR PIR; S08495; RSVF32.
DR PIR; S08444; S08444.
DR Mendel; 5435; VICfa; rpl32;1.
DR InterPro: IPR002677; Ribosomal_L32p.
DR Pfam; PF01783; Ribosomal_L32p; 1.
KW Ribosomal protein; Chloroplast.
SQ SEQUENCE 48 AA; 5407 MW; 146372D6C9BC34DB CRC64;

Query Match 12.3%; Score 34; DB 1; Length 48;
Best Local Similarity 42.9%; Pred. No. 6.4e+02;
Matches 9; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 26 YEAPVLHFSEALHIIAGTPEV 46
DB 25 YKAALKAFSLADSLTGTSKV 45
```

```
RESULT 12
INS_BALBO STANDARD; PRT; 51 AA.
ID INS_BALBO
AC P01314;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
GN INS.
OS Balaenoptera borealis (Sei whale).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9768;
RN [1]
RP SEQUENCE.
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin."
RL Nature 181:1468-1469(1958).
CC -1- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -1- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01582; INWHIS.
DR HSSP; P01315; 9INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULINA.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 12.3%; Score 34; DB 1; Length 51;
Best Local Similarity 32.0%; Pred. No. 6.8e+02;
Matches 8; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

QY 32 HFSEALHIIAG-----TPEVHAVK 50
DB 10 HLVEALYLCGERGFYTPKAGIVE 34

RESULT 13
INS_BALPH STANDARD; PRT; 51 AA.
ID INS_BALPH
AC P01312;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Insulin.
GN INS.
OS Balaenoptera physalus (Finback whale) (Common rorqual), and
OS Physeter catodon (Sperm whale) (Physeter macrocephalus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Cetartiodactyla; Cetacea; Mysticeti;
OC Balaenopteridae; Balaenoptera.
OX NCBI_TaxID=9770, 9755;
RN [1]
RP PARTIAL SEQUENCE.
RC SPECIES=B.physalus;
```

```

RA Hana H., Titani K., Sakaki S., Narita K.;
RT "The amino acid sequence in fin-whale insulin.";
RN J. Biochem. 56:285-293(1964).
RP [2]
RC SPECIES=P.catodon;
RA Ishihara Y., Saito T., Ito Y., Fujino M.;
RT "Structure of sperm- and sei-whale insulins and their breakdown by
RT whale pepsin.";
RL Nature 181:1468-1469(1958).
RN [3]
RP SEQUENCE.
RC SPECIES=P.catodon;
RA Harris J.I., Sanger F., Naughton M.A.;
RT "Species differences in insulin.";
RL Arch. Biochem. Biophys. 65:427-438(1956).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC PIR; A91918; INWHF.
DR HSSP; P01315; 9INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT SEQUENCE 51 AA; 5766 MW; 9007B514691A7CDD CRC64;

Query Match 12.3%; Score 34; DB 1; Length 51;
Best Local Similarity 32.0%; Pred. No. 6.8e+02;
Matches 8; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

Qy 32 HFSEALHIIAG-----TPEVHAVK 50
Db 10 HLVEALYLCVGERGFFYTPKAGIVE 34

RESULT 14
INS_CAMDR
ID INS_CAMDR STANDARD; PRT; 51 AA.
AC P01320;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Camelus dromedarius (Dromedary) (Arabian camel).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
OX NCBI_TaxID=9838;
RN [1]
RP SEQUENCE.
RA Danho W.O.;
RT "The isolation and characterization of insulin of camel (Camelus
RT dromedarius).";
RL J. Fac. Med. Baghdad 14:16-28(1972).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND

```

```

CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR HSSP; P01317; 2INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41
FT SEQUENCE 51 AA; 5693 MW; 901E88BA085A7DDD CRC64;

Query Match 12.3%; Score 34; DB 1; Length 51;
Best Local Similarity 32.0%; Pred. No. 6.8e+02;
Matches 8; Conservative 5; Mismatches 6; Indels 6; Gaps 1;

Qy 32 HFSEALHIIAG-----TPEVHAVK 50
Db 10 HLVEALYLCVGERGFFYTPKAGIVE 34

RESULT 15
INS_CAPHI
ID INS_CAPHI STANDARD; PRT; 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RL "Species variation in the amino acid sequence of insulin.";
CC Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: INSULIN DECREASES BLOOD GLUCOSE CONCENTRATION. IT
CC INCREASES CELL PERMEABILITY TO MONOSACCHARIDES, AMINO ACIDS AND
CC FATTY ACIDS. IT ACCELERATES GLYCOLYSIS, THE PENTOSE PHOSPHATE
CC CYCLE, AND GLYCOGEN SYNTHESIS IN LIVER.
CC -!- SUBUNIT: HETERODIMER OF A B CHAIN AND AN A CHAIN LINKED BY TWO
CC DISULFIDE BONDS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
DR PIR; A01586; INGT.
DR HSSP; P01315; 9INS.
DR InterPro: IPR000739; Insulin_IGF_relaxin.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00276; INSULIN.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:00:18 ; Search time 23.96 Seconds
(without alignments)
382.668 Million cell updates/sec

Title: US-09-696-169A-20

Perfect score: 277

Sequence: 1 QLVPKLDEVYNAAYNAADHA.....SEALHIITAGTPEVHAVKPGA 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 33061

Minimum DB seq length: 0

Maximum DB seq length: 53

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mbc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	15.9	42	10 Q9LEJ5	Q9lej5 glycine max
2	44	15.9	45	10 Q9LEJ39	Q9lej39 glycine max
3	41	14.8	44	2 Q9APO5	Q9apq5 uncultured
4	41	14.8	48	2 Q9RID3	Q9rid3 yersinia pe
5	39	14.1	53	5 Q9VQ65	Q9vq65 drosophila
6	38	13.7	31	10 Q9LEJ3	Q9lej3 glycine max
7	38	13.7	37	6 Q18822	Q18822 sus scrofa
8	38	13.7	40	6 Q9TTD6	Q9tttd6 sus scrofa
9	38	13.7	48	2 Q02801	Q02801 streptomyc
10	38	13.7	50	5 Q27271	Q27271 plasmodium
11	38	13.7	53	16 Q92ZX1	Q92zx1 rhizobium m
12	37	13.4	42	2 Q9X8F1	Q9x8f1 streptomyc
13	37	13.4	46	4 Q15456	Q15456 homo sapien
14	37	13.4	50	5 Q26209	Q26209 plasmodium
15	37	13.4	50	5 Q26213	Q26213 plasmodium
16	36.5	13.2	39	16 Q9KDC1	Q9kdc1 bacillus ha

17	36.5	13.2	40	2 Q45678	Q45678 bacillus su
18	36	13.0	48	16 Q92L87	Q92l87 helicobacte
19	36	13.0	50	2 Q939R3	Q939r3 bacillus th
20	36	13.0	51	2 Q9R636	Q9r636 rhodobacter
21	36	13.0	53	12 Q56128	Q56128 porcine cir
22	36	13.0	53	12 Q91861	Q91861 porcine cir
23	35.5	12.8	45	16 Q34405	Q34405 bacillus su
24	35.5	12.8	50	2 Q9KGZ4	Q9kgz4 clostridium
25	35.5	12.8	51	2 Q34277	Q34277 yersinia pe
26	35.5	12.8	53	2 Q93LH2	Q93lh2 sulfitobact
27	35	12.6	29	4 Q96EJ1	Q96ej1 homo sapien
28	35	12.6	36	17 Q26422	Q26422 methanother
29	35	12.6	37	5 Q9NL22	Q9nl22 leishmania
30	35	12.6	42	16 Q97SZ7	Q97sz7 streptococ
31	35	12.6	46	9 Q9TJ08	Q9tj08 bacterioph
32	35	12.6	46	10 Q49699	Q49699 arabidopsis
33	35	12.6	46	16 Q9I627	Q9i627 pseudomonas
34	35	12.6	53	2 Q52116	Q52116 brucella ab
35	35	12.6	53	2 Q52635	Q52635 brucella me
36	34.5	12.5	32	5 Q95NY7	Q95ny7 drosophila
37	34.5	12.5	45	5 Q9WIR2	Q9wir2 drosophila
38	34.5	12.5	50	2 Q9ZHW6	Q9zhw6 enterococcu
39	34.5	12.5	53	16 Q9A574	Q9a574 caulobacter
40	34	12.3	18	6 Q9GMH1	Q9gmh1 macaca mula
41	34	12.3	20	6 Q95MK3	Q95mk3 ateles belz
42	34	12.3	33	5 Q9NL21	Q9nl21 leishmania
43	34	12.3	39	10 Q9FEY1	Q9fev1 heterocapsa
44	34	12.3	39	16 Q9KVK5	Q9kvk5 vibrio chol
45	34	12.3	40	5 Q9I814	Q9i814 drosophila

ALIGNMENTS

RESULT 1
Q9LEJ5 PRELIMINARY; PRT; 42 AA.
ID Q9LEJ5
AC Q9LEJ5;
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-MAR-2001 (TREMREL. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodkin L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293440; CAB96762.1; -;
FT NON_TER
SQ SEQUENCE 42 AA; 4935 MW; 2327B258B8A2CFA5 CRC64;

Query Match 15.9%; Score 44; DB 10; Length 42;
Best Local Similarity 31.6%; Pred. No. 1.6e+02;
Matches 12; Conservative 6; Mismatches 12; Indels 8; Gaps 2;
QY 14 YNAADHAPE-DKYEAFVLFSEALHIITAGTPEVHAVK 50
DB 10 YEKVDHFAPEPTKYKDLVVKLTKNV-----EALHVE 40

```

RESULT 2
ID Q9LE39 PRELIMINARY; PRT; 45 AA.
AC Q9LE39;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).
GN MSG.
OS Glycine max (Soybean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
OX NCBI_TaxID=3847;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RA Vodka L.O.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CLARK;
RX MEDLINE=20044090; PubMed=10579489;
RA Stromvik M.V., Sundaraman V.P., Vodka L.O.;
RT "A novel promoter that is active in a complex developmental pattern
RT with and without its proximal 650 base pairs.";
RL Plant Mol. Biol. 41:217-231(1999).
DR EMBL; AJ293442; CAB96764.1; -.
DR EMBL; AJ293443; CAB96760.1; -.
FT NON_TER 1
SQ SEQUENCE 45 AA; 5236 MW; 2327E24CEA15458E CRC64;

Query Match 15.9%; Score 44; DB 10; Length 45;
Best Local Similarity 31.6%; Pred. No. 1.7e+02;
Matches 12; Conservative 6; Mismatches 12; Indels 8; Gaps 2;

QY 14 YNAADHAAPE-DKYAFVLFHSEALHIIAGTPEVHAVK 50
| | | | | | | | | | | | | | | | | | | | |
DB 13 YEKVDHTAPEPTKYKDLVVKLTKNV-----EALHVE 43
| | | | | | | | | | | | | | | | | | | | |

RESULT 3
ID Q9APQ5 PRELIMINARY; PRT; 44 AA.
AC Q9APQ5;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE ELSG (FRAGMENT).
GN ELSG.
OS uncultured bacterium pCosHEL.
OC Bacteria; environmental samples.
OX NCBI_TaxID=143796;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20575196; PubMed=11133432;
RA Entcheva P., Liebl W., Johann A., Hartsch T., Streit W.R.;
RT "Direct cloning from enrichment cultures, a reliable strategy for
RT isolation of complete operons and genes from microbial consortia.";
RL Appl. Environ. Microbiol. 67:89-99(2001).
DR EMBL; AF250771; AAG60565.1; -.
FT NON_TER 44
SQ SEQUENCE 44 AA; 5262 MW; 6C73B171FDE4F678 CRC64;

Query Match 14.8%; Score 41; DB 2; Length 44;
Best Local Similarity 37.5%; Pred. No. 4e+02;
Matches 9; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 7 DEVYNAAYNAADHAAPEDKYAEFV 30

```

```

Db 18 DEYPAVFAIDRAQHKVILETFI 41
| | | | | | | | | | | | | | | | | | | | |

RESULT 4
ID Q9RID3 PRELIMINARY; PRT; 48 AA.
AC Q9RID3;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 5.4 KDA PROTEIN.
GN YMT1.42AC.
OS Yersinia pestis.
OG Plasmid pMT1.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Baker S.G., Mungall K.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 BIOVAR ORIENTALIS;
RA Karlyshev A.V., Wren B.W.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117211; CAB62369.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 48 AA; 5417 MW; BFD2188AF24BED48 CRC64;

Query Match 14.8%; Score 41; DB 2; Length 48;
Best Local Similarity 31.4%; Pred. No. 4.4e+02;
Matches 11; Conservative 5; Mismatches 19; Indels 0; Gaps 0;

QY 6 LDEVYNAAYNAADHAAPEDKYAEFVLFHSEALHII 40
| | | | | | | | | | | | | | | | | | | | |
DB 1 MSEFYSGRAAAVADQAGNEDERAGPVFNLPPLIIV 35
| | | | | | | | | | | | | | | | | | | | |

RESULT 5
ID Q9VQ65 PRELIMINARY; PRT; 53 AA.
AC Q9VQ65;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DE 01-MAY-2000 (TRENBLrel. 13, Last annotation update)
DE CG15353 PROTEIN.
GN CG15353.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=107311132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

```

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreria S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.W., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003584; AAF51316.1; -;
 DR FlyBase; FBgn0040718; CG15353.
 SQ SEQUENCE 53 AA; 5450 MW; 663595A2F81D2E9E CRC64;

Query Match 14.1%; Score 39; DB 5; Length 53;
 Best Local Similarity 45.0%; Pred. No. 9e+02;

Matches 9; Conservative 4; Mismatches 5; Indels 2; Gaps 1;

QY 3 VPKLDEYVNAAYNAADHAP 22

II : | | | | : | |

Db 31 VPQF-VYSAGYPAGVYASP 48

RESULT 6

ID Q9LEJ3 PRELIMINARY; PRT; 31 AA.

AC Q9LEJ3;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MAJOR LATEX PROTEIN HOMOLOGUE (FRAGMENT).

GN MSG.

OS Glycine max (Soybean).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

OX NCBI_TaxID=3847;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. WILLIAMS 82;

RC TISSUE=SHOOT APEX, 9-10 DAY OLD ETIOLATED SEEDLINGS;

RA Vodkin L.O.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV. WILLIAMS 82;

RC TISSUE=SHOOT APEX, 9-10 DAY OLD ETIOLATED SEEDLINGS;

RX MEDLINE=20044090; PubMed=10579489;

RA Stromvik M.V., Sundaraman V.P., Vodkin L.O.;

RT "A novel promoter that is active in a complex developmental pattern

with and without its proximal 650 base pairs.";

PL Plant Mol. Biol. 41:217-231(1999).

DR EMBL; AJ293444; CAB96765.1; -;
 FT NON_TER 1
 SQ SEQUENCE 31 AA; 3530 MW; 4FE1EB734C2C5B4E CRC64;

Query Match 13.7%; Score 38; DB 10; Length 31;
 Best Local Similarity 32.4%; Pred. No. 6.5e+02;

Matches 11; Conservative 6; Mismatches 9; Indels 8; Gaps 2;

QY 18 DHAPE-DKYEAFVLHSEALHIIAGTPEVHAVK 50

II ||| || : : : | | |

Db 3 DHTAPEPTYKDLVWLTKNV-----EAHLVE 29

RESULT 7

OL8822 PRELIMINARY; PRT; 37 AA.

AC OL8822;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE AGOUTI (FRAGMENT).

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=YORKSHIRE;

RX MEDLINE=98398091; PubMed=9730323;

RA Wang Y., Westby C.A., Johansen M., Marshall D.M., Granholm N.;

RT "Isolation, cloning, and sequencing of porcine agouti exon 2

(PorAex2).";

RL Pigment Cell Res. 11:155-157(1998).

DR EMBL; AF018166; AAC34732.1; -;

FT NON_TER 1

FT NON_TER 37

SQ SEQUENCE 37 AA; 4100 MW; 5555E9CE37DDF6DA CRC64;

Query Match 13.7%; Score 38; DB 6; Length 37;

Best Local Similarity 50.0%; Pred. No. 8e+02;

Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 14 YNAADHAAPEDK 25

: | : | | | |

Db 10 FTASSHLAPEEK 21

RESULT 8

Q9TTD6

ID Q9TTD6 PRELIMINARY; PRT; 40 AA.

AC Q9TTD6;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE AGOUTI (FRAGMENT).

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20292452; PubMed=10834600;

RA Kim K.S., Mendez E.A., Marklund S., Clutter A.C., Pomp D.,

RA Rothschild M.F.;

RT "Linkage mapping of the porcine Agouti gene.";

RL J. Anim. Sci. 78:1395-1396(2000).

DR EMBL; AF133261; AAF22158.1; -;

FT NON_TER 1

FT NON_TER 40

SQ SEQUENCE 40 AA; 4447 MW; 7239F4E34CAE419E CRC64;

Query Match 13.7%; Score 38; DB 6; Length 40;
Best Local Similarity 50.0%; Pred. No. 8.e+02;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 14 YNAADHAAPEDK 25
Db 5 FTASHLAPEK 16

RESULT 9

ID Q02801 PRELIMINARY; PRT; 48 AA.
AC Q02801;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)
DE HYPOTHETICAL PROTEIN IN PTH4 PROMOTER (FRAGMENT).
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=93163057; PubMed=7679386;
RA Tan H., Chater K.F.;
RT "Two developmentally controlled promoters of Streptomyces coelicolor
A3(2) that resemble the major class of motility-related promoters in
other bacteria."
RT J. Bacteriol. 175:933-940(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bruton C.J.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; X68791; CAA48686.1; -;
KW Hypothetical protein.
FT NON_TER 48
SQ SEQUENCE 48 AA; 4815 MW; 0F2D3913881B1735 CRC64;

Query Match 13.7%; Score 38; DB 2; Length 48;
Best Local Similarity 34.6%; Pred. No. 1.1e+03;
Matches 9; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 28 AFVLHFSEALHIIAGTPEVHAVKGA 53
Db 11 AALLSASACGLTSGSPMVDVPEGS 36

RESULT 10

ID Q27271 PRELIMINARY; PRT; 50 AA.
AC Q27271;
DT 01-NOV-1996 (TREMELrel. 01, Created)
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMELrel. 08, Last annotation update)
DE RHOPTRY PROTEIN (FRAGMENT).
OS Plasmodium berghei yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5862;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=95364832; PubMed=7637695;
RA Borre M.B., Owen C.A., Keen J.K., Sinha K.A., Holder A.A.;
RT "Multiple genes code for high-molecular-mass rhostry proteins of
Plasmodium yoelii."
RL Mol. Biochem. Parasitol. 70:149-155(1995).
DR EMBL; U19985; AAC46620.1; -;
DR EMBL; U19979; AAC46614.1; -;
FT NON_TER 1

FT NON_TER 50
SQ SEQUENCE 50 AA; 6042 MW; 33B9D1EA07A4265E CRC64;

Query Match 13.7%; Score 38; DB 5; Length 50;
Best Local Similarity 24.5%; Pred. No. 1.1e+03;
Matches 12; Conservative 13; Mismatches 22; Indels 2; Gaps 2;

QY 2 LVPKLDEVYNAAYNAADHAAPEDKYEAFVLHFSEALHIIAGTPEVHAVK 50
Db 1 LKNIHV-NKAKSYLDY-VKNEFDRIYTHFKNLNTVNDKFKVEYLK 47

RESULT 11

Q922X1 PRELIMINARY; PRT; 53 AA.
AC Q922X1;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SMA0625.
GN SMA0625.
OS Rhizobium melliloti (Sinorhizobium melliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
Sinorhizobium melliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007225; AAK64985.1; -;
KW Hypothetical protein; Plasmid; Complete proteome.
SQ SEQUENCE 53 AA; 5568 MW; 40316BEA492AB2DB CRC64;

Query Match 13.7%; Score 38; DB 16; Length 53;
Best Local Similarity 41.2%; Pred. No. 1.2e+03;
Matches 7; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 28 AFVLHFSEALHIIAGTP 44
Db 18 AFMLFLQSVVHAFAGQP 34

RESULT 12

Q9X8F1 PRELIMINARY; PRT; 42 AA.
AC Q9X8F1;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE HYPOTHETICAL 4.2 KDA PROTEIN.
GN SCE39.28.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Oliver K., Harris D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBSJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL049573; CAB40336.1; -.
KW Hypothetical protein.
SQ SEQUENCE 42 AA; 4168 MW; 999E9AABE3EA1500 CRC64;

Query Match 13.4%; Score 37; DB 2; Length 42;
Best Local Similarity 58.3%; Pred. No. 1.3e+03;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 42 GTPVEHVAVKPGA 53
DQ 8 GVPSRHAVREGA 19

RESULT 13
Q15456 PRELIMINARY; PRT; 46 AA.
AC Q15456;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROTEIN-SERINE/THREONINE KINASE (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94100173; PubMed=8274451;
RA Schultz S.J., Nigg E.A.;
RT "Identification of 21 novel human protein kinases, including 3 members
RT of a family related to the cell cycle regulator nima of Aspergillus
RT nidulans.";
RL Cell Growth Differ. 4:821-830(1993).
DR EMBL; Z25435; CAA80922.1; -.
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; pkinase.1
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Kinase; Transferase.
FT NON_TER 1
FT NON_TER 46
SQ SEQUENCE 46 AA; 5140 MW; 66C551D29D7382D9 CRC64;

Query Match 13.4%; Score 37; DB 4; Length 46;
Best Local Similarity 34.5%; Pred. No. 1.4e+03;
Matches 10; Conservative 2; Mismatches 17; Indels 0; Gaps 0;

QY 14 YNAADHAAPEDKYEAFVLHFSALHIAG 42
DQ 5 FGSASHVADNDITPYLVSRFYRAPEIIIG 33

RESULT 14
Q26209 PRELIMINARY; PRT; 50 AA.
AC Q26209;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RHOPTRY PROTEIN (FRAGMENT).
```

```
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=YM;
RX MEDLINE=95364832; PubMed=7637695;
RA Borre M.B., Owen C.A., Keen J.K., Sinha K.A., Holder A.A.;
RT "Multiple genes code for high-molecular-mass rhostry proteins of
RT Plasmodium yoelii.";
RL MOL. Biochem. Parasitol. 70:149-155(1995).
DR EMBL; U19980; AAC46615.1; -.
FT NON_TER 1
FT NON_TER 50
SQ SEQUENCE 50 AA; 6023 MW; C378C78A42AA43C2 CRC64;
```

Query Match 13.4%; Score 37; DB 5; Length 50;
Best Local Similarity 25.6%; Pred. No. 1.5e+03;
Matches 10; Conservative 12; Mismatches 15; Indels 2; Gaps 2;

```
QY 2 LVPKLDVYNAAYNAADHAAPEDKYEAFVLHFSALHII 40
DQ 1 LKXSIENV-NKAKSYLDY-IKENEFDRIVTHFKKKLNTV 37
```

```
RESULT 15
Q26213 PRELIMINARY; PRT; 50 AA.
AC Q26213;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RHOPTRY PROTEIN (FRAGMENT).
OS Plasmodium yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5861;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=YM;
RX MEDLINE=95364832; PubMed=7637695;
RA Borre M.B., Owen C.A., Keen J.K., Sinha K.A., Holder A.A.;
RT "Multiple genes code for high-molecular-mass rhostry proteins of
RT Plasmodium yoelii.";
RL MOL. Biochem. Parasitol. 70:149-155(1995).
DR EMBL; U19987; AAC46622.1; -.
FT NON_TER 1
FT NON_TER 50
SQ SEQUENCE 50 AA; 6050 MW; F801F78A42AA5848 CRC64;
```

Query Match 13.4%; Score 37; DB 5; Length 50;
Best Local Similarity 25.6%; Pred. No. 1.5e+03;
Matches 10; Conservative 12; Mismatches 15; Indels 2; Gaps 2;

```
QY 2 LVPKLDVYNAAYNAADHAAPEDKYEAFVLHFSALHII 40
DQ 1 LKXSIENV-NKAKSYLDY-IKENEFDRIVTHFKKKLNTV 37
```

Search completed: July 11, 2002, 11:03:43
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:56:43 ; Search time 29.64 seconds

(without alignments)
198.614 Million cell updates/sec

Title: US-09-696-169A-20

Perfect score: 277

Sequence: 1 OLVPKLDEVYNAAYNAADHA.....SEALHIAGTPEVHAVKPGA 53

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 362426

Minimum DB seq length: 0

Maximum DB seq length: 53

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/hold-geneseg/genesegp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	70	25.3	20 17 AAR89393	Kentucky Blue Grass
2	65	23.5	20 16 AAR71517	LPIX-10, peptide f
3	57	20.6	20 16 AAR71516	LPIX-9, peptide fr
4	56	20.2	20 16 AAR71518	LPIX-11, peptide f
5	54	19.5	20 16 AAR71519	LPIX-12, peptide f
6	49	17.7	20 17 AAR89392	Kentucky Blue Grass
7	48	17.3	20 16 AAR71520	LPIX-13, peptide f
8	47	17.0	20 17 AAR89391	Kentucky Blue Grass
9	44.5	16.1	53 22 AAG93042	C glutamicum prote
10	44	15.9	12 19 AAW76473	Graminae pollen al
11	44	15.9	12 19 AAW76485	Graminae pollen al

12	44	15.9	29 18 AAW10845	MAB anti-HBSag bin
13	44	15.9	33 18 AAW10860	MAB anti-HBSag bin
14	43.5	15.7	36 16 AAR80961	Integrin alpha-6a
15	43	15.5	53 22 ABB28520	Peptide #1171 enco
16	43	15.5	53 22 ABB33705	Peptide #1211 enco
17	43	15.5	53 22 ABB19154	Protein #1153 enco
18	43	15.5	53 22 AAM54473	Human brain expro
19	43	15.5	53 22 AAM66878	Human bone marrow
20	43	15.5	53 22 AAM14740	Peptide #1174 enco
21	43	15.5	53 22 AAM27163	Peptide #1200 enco
22	43	15.5	53 22 AAM02462	Peptide #1144 enco
23	41	14.8	20 16 AAR71513	LPIX-6, peptide fr
24	41	14.8	27 22 ABB27510	Novel human diagno
25	41	14.8	53 20 AAY42432	Amino acid residue
26	40	14.4	12 19 AAW76486	Graminae pollen al
27	40	14.4	32 19 AAW73566	Insulin 1 B-chain.
28	40	14.4	51 19 AAW79410	Staphylococcus aur
29	40	14.4	51 21 AAB32429	Human secreted pro
30	39	14.1	12 19 AAW76484	Graminae pollen al
31	39	14.1	20 16 AAR71512	LPIX-5, peptide fr
32	39	14.1	20 17 AAR89394	Kentucky Blue Gras
33	39	14.1	34 19 AAW66394	Cationic peptide a
34	39	14.1	34 21 AAY91693	Cationic peptide A
35	39	14.1	35 22 AAB51195	E. coli AMP gene a
36	39	14.1	39 22 AAB99854	Physcomitrella pat
37	39	14.1	40 22 ABB41608	Peptide #9114 enco
38	39	14.1	40 22 ABB25434	Protein #7433 enco
39	39	14.1	40 22 AAM62479	Human brain expro
40	39	14.1	40 22 AAM75288	Human bone marrow
41	39	14.1	40 22 AAM20559	Peptide #6993 enco
42	39	14.1	40 22 AAM35401	Peptide #9438 enco
43	39	14.1	41 21 AAB56544	Human prostate can
44	39	14.1	41 22 AAU06521	Human CYP2B6 prote
45	39	14.1	41 22 AAU06522	Human CYP2B6 prote

ALIGNMENTS

RESULT 1
AAR89393
ID AAR89393 standard; peptide; 20 AA.
XX
AC AAR89393;
XX
DT 17-SEP-1996 (first entry)
DE Kentucky Blue Grass pollen allergen rKBG60 resides 129-148.
XX

KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
KW allergic; allergy; prevention; development; reaction; grass;
KW antibody; anergise; producing cell; diagnosis; human; clone 60;
KW immunological response; B-cell; hlgG; hlgE; T-cell.
XX
OS Poa pratensis.
XX
PN WO9603106-A2.
XX
PD 08-FEB-1996.
XX
PF 26-JUL-1995; 95WO-CA00439.
XX
PR 26-JUL-1994; 94US-0280455.
XX
PA (UYMA-) UNIV MANITOBA.
XX
PI Mohapatra SS;
XX
DR WPI; 1996-116753/12.
XX
PT Peptide including at least 1 human antigenic determinant of Kentucky
Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,

PT used to desensitise allergic individuals, prevent grass pollen
 XX allergy and deplete allergen-specific antibodies

PS Example 5; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide,
 CC comprising residues 129-148 of the recombinant Kentucky Blue Grass
 CC pollen allergen rkg60, which is from the Poa p ix gp. of grass
 CC pollen allergens. A compsn. comprising the peptide and a carrier
 CC can be used to desensitise allergic individuals, and prevent
 CC allergic individuals from developing an allergic reaction to grass
 CC pollen. The compsn. may also be used to deplete allergen specific
 CC antibodies from an individual, and anergise allergen specific
 CC antibody producing cells. The peptide can be used to diagnose grass
 CC pollen allergy, and has the following human immunological response
 CC profile, B-cell (hlg (-)/hlgE (-)), T-cell (-).

SQ Sequence 20 AA;

Query Match 25.3%; Score 70; DB 17; Length 20;

Best Local Similarity 83.3%; Pred. No. 0.0073;

Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 34 SEALHRIAGTPEVHAVKP 51

Db 2 sealriagtlevhgvp 19

RESULT 2

AAR71517

ID AAR71517 standard; Protein; 20 AA.

AC AAR71517;

DT 02-NOV-1995 (first entry)

DE LPIX-10, peptide fragment of Lol pv protein allergen.

KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;

XX sensitivity; ryegrass pollen allergen.

OS Lolium perenne.

PN W09506728-A.

PD 09-MAR-1995.

PF 05-AUG-1994; 94WO-US09024.

PR 13-AUG-1993; 93US-0106016.

XX (IMMU-) IMMULOGIC PHARM CORP.

PI Griffith IJ, Kuo M, Luqman M;

DR WPI; 1995-115444/15.

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)

PT and DNA - for treating sensitivity to rye-grass pollen allergen

PT or an immunologically cross-reactive allergen.

PS Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 23.5%; Score 65; DB 16; Length 20;

Best Local Similarity 60.0%; Pred. No. 0.036;

Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 YNAYNAADHAAPEDKYEF 29

Db 1 ydvaykaegatpeakydaf 20

RESULT 3

AAR71516

ID AAR71516 standard; Protein; 20 AA.

AC AAR71516;

DT 02-NOV-1995 (first entry)

DE LPIX-9, peptide fragment of Lol pv protein allergen.

KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;

XX sensitivity; ryegrass pollen allergen.

OS Lolium perenne.

PN W09506728-A.

PD 09-MAR-1995.

PF 05-AUG-1994; 94WO-US09024.

PR 13-AUG-1993; 93US-0106016.

XX (IMMU-) IMMULOGIC PHARM CORP.

PI Griffith IJ, Kuo M, Luqman M;

DR WPI; 1995-115444/15.

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)

PT and DNA - for treating sensitivity to rye-grass pollen allergen

PT or an immunologically cross-reactive allergen.

PS Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).

SQ Sequence 20 AA;

Query Match 20.6%; Score 57; DB 16; Length 20;

Best Local Similarity 58.8%; Pred. No. 0.47;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 LVPKLDENVNAAYNAAD 18

Db 3 lipkldtaydvaykaae 19

RESULT 4

AAR71518

ID AAR71518 standard; Protein; 20 AA.

AC AAR71518;

DT 02-NOV-1995 (first entry)

XX DE LP1X-11, peptide fragment of Lol pv protein allergen.
 XX KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 XX PT sensitivity; ryegrass pollen allergen.
 XX OS Lolium perenne.
 XX PN WO9506728-A.
 XX PD 09-MAR-1995.
 XX PF 05-AUG-1994; 94WO-US09024.
 XX PR 13-AUG-1993; 93US-0106016.
 XX PA (IMMU-) IMMULOGIC PHARM CORP.
 XX PI Griffith IJ; Kuo M, Lugman M;
 XX DR WPI; 1995-115444/15.
 XX PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 XX PT and DNA - for treating sensitivity to rye-grass pollen allergen
 XX PT or an immunologically cross-reactive allergen.
 XX PS Claim 1; Fig 2; 110pp; English.
 XX CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).
 XX SQ Sequence 20 AA;

Query Match 20.2%; Score 56; DB 16; Length 20;
 Best Local Similarity 55.0%; Pred. No. 0.65;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 20 AAPDKYAEFVLHFSEALHI 39
 I - - - - -
 Db 1 atpeakydafvtaltealrv 20

RESULT 5
 AAR71519
 ID AAR71519 standard; Protein; 20 AA.
 XX AC AAR71519;
 XX DT 02-NOV-1995 (first entry)
 XX DE LP1X-12, peptide fragment of Lol pv protein allergen.
 XX KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 XX PT sensitivity; ryegrass pollen allergen.
 XX OS Lolium perenne.
 XX PN WO9506728-A.
 XX PD 09-MAR-1995.
 XX PF 05-AUG-1994; 94WO-US09024.
 XX PR 13-AUG-1993; 93US-0106016.
 XX PA (IMMU-) IMMULOGIC PHARM CORP.

PI Griffith IJ, Kuo M, Lugman M;
 XX WPI; 1995-115444/15.
 XX PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 XX PT and DNA - for treating sensitivity to rye-grass pollen allergen
 XX PT or an immunologically cross-reactive allergen.
 XX PS Claim 1; Fig 2; 110pp; English.
 XX CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).
 XX SQ Sequence 20 AA;

Query Match 19.5%; Score 54; DB 16; Length 20;
 Best Local Similarity 60.0%; Pred. No. 1.2;
 Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 30 VLHFSEALHIAGTPEVHAV 49
 I - - - - -
 Db 1 vtaalrviagalevhav 20

RESULT 6
 AAR89392
 ID AAR89392 standard; peptide; 20 AA.
 XX AC AAR89392;
 XX DT 17-SEP-1996 (first entry)
 XX DE Kentucky Blue Grass pollen allergen rKBG60 resides 109-128.
 XX KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
 KW allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
 KW allergic; allergy; prevention; development; reaction; grass;
 KW antibody; anergise; producing cell; diagnosis; human; clone 60;
 XX immunological response; B-cell; hlgg; hlgg; T-cell.
 OS Poa pratensis.
 XX PN WO9603106-A2.
 XX PD 08-FEB-1996.
 XX PF 26-JUL-1995; 95WO-CA00439.
 XX PR 26-JUL-1994; 94US-0280455.
 XX PA (UYMA-) UNIV MANITOBA.
 XX PI Mohapatra SS;
 XX DR WPI; 1996-116753/12.
 XX PT Peptide including at least 1 human antigenic determinant of Kentucky
 PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,
 PT used to desensitise allergic individuals, prevent grass pollen
 PT allergy and deplete allergen-specific antibodies
 XX PS Claim 8; Page 29; 59pp; English.
 XX CC The present peptide is a human antigenic determinant contg. peptide,
 CC comprising residues 109-128 of the recombinant Kentucky Blue Grass
 CC pollen allergen rKBG60, which is from the Poa p IX gp. of grass
 CC pollen allergens. A compsn. comprising the peptide and a carrier

CC can be used to desensitise allergic individuals, and prevent
 CC allergic individuals from developing an allergic reaction to grass
 CC pollen. The compsn. may also be used to deplete allergen specific
 CC antibodies from an individual, and anergise allergen specific
 CC antibody producing cells. The peptide can be used to diagnose grass
 CC pollen allergy, and has the following human immunological response
 CC profile, B-cell (hlg (-)/higE (-)), T-cell (+).

XX Sequence 20 AA;

Query Match 17.7%; Score 49; DB 17; Length 20;
 Best Local Similarity 50.0%; Pred. No. 6.2;
 Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 13 AYNRAADHAAPEDKYEAFV 30
 || : | : | | | : | : |
 Db 1 ayksaegatpeakyddv 18

RESULT 7

AAR71520
 ID AAR71520 standard; Protein; 20 AA.

XX AC AAR71520;

XX DT 02-NOV-1995 (first entry)

XX DE LPIX-13, peptide fragment of Lol pv protein allergen.

XX Lollum perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 KW sensitivity; ryegrass pollen allergen.

XX OS Lollum perenne.

XX PN W09506728-A.

XX PD 09-MAR-1995.

XX PF 05-AUG-1994; 94WO-US09024.

XX PR 13-AUG-1993; 93US-0106016.

XX PA (IMMU-) IMMULOGIC PHARM CORP.

XX PI Griffith IJ, Kuo M, Luqman M;

XX DR WPI; 1995-115444/15.

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 PT and DNA - for treating sensitivity to rye-grass pollen allergen
 PT or an immunologically cross-reactive allergen.

XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 17.3%; Score 48; DB 16; Length 20;
 Best Local Similarity 83.3%; Pred. No. 8.5;
 Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51

Db 1 iagalevhavkp 12

RESULT 8

AAR89391

ID AAR89391 standard; peptide; 20 AA.

XX AC AAR89391;

XX DT 17-SEP-1996 (first entry)

XX DE Kentucky Blue Grass pollen allergen rKBG60 residues 99-118.

XX Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
 KW allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
 KW allergic; allergy; prevention; development; reaction; grass;
 KW antibody; anergise; producing cell; diagnosis; human; clone 60;
 KW immunological response; B-cell; hlgg; hlgE; T-cell.

XX OS Poa pratensis.

XX PN W09603106-A2.

XX PD 08-FEB-1996.

XX PF 26-JUL-1995; 95WO-CA00439.

XX PR 26-JUL-1994; 94US-0280455.

XX PA (UYMA-) UNIV MANITOBA.

XX PI Mohapatra SS;

XX DR WPI; 1996-116753/12.

XX Peptide including at least 1 human antigenic determinant of Kentucky
 PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,
 PT used to desensitise allergic individuals, prevent grass pollen
 PT allergy and deplete allergen-specific antibodies

XX Claim 7; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide,
 CC comprising residues 99-118 of the recombinant Kentucky Blue Grass
 CC pollen allergen rKBG60, which is from the Poa p IX gp. of grass
 CC pollen allergens. A compsn. comprising the peptide and a carrier
 CC can be used to desensitise allergic individuals, and prevent
 CC allergic individuals from developing an allergic reaction to grass
 CC pollen. The compsn. may also be used to deplete allergen specific
 CC antibodies from an individual, and anergise allergen specific
 CC antibody producing cells. The peptide can be used to diagnose grass
 CC pollen allergy, and has the following human immunological response
 CC profile, B-cell (hlg (-)/higE (+)), T-cell (+).

XX Sequence 20 AA;

Query Match 17.0%; Score 47; DB 17; Length 20;

Best Local Similarity 50.0%; Pred. No. 12;

Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

QY 5 KLDEVYNAAYNAADHAAP 22

Db 3 kldaayklayksaegatp 20

RESULT 9

AAG93042

ID AAG93042 standard; Protein; 53 AA.

XX AC AAG93042;

XX DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 6796.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR N-PSDB; AAH68261.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 6796; 246pp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from Coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 53 AA;

Query Match 16.1%; Score 44.5; DB 22; Length 53;
Best Local Similarity 35.7%; Pred. NO. 89;
Matches 15; Conservative 4; Mismatches 16; Indels 7; Gaps 1;

QY 15 NAADHAAPEDKVEA-----FVLHFSALHTIAGTPEVHAV 49
DB 2 nlvdhdpensgegragvmlngtlkgtdmilitagdpdvav 43

RESULT 10
AAW76473
ID AAW76473 standard; peptide; 12 AA.
XX
AC AAW76473;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #25.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX

Query Match 15.9%; Score 44; DB 19; Length 12;
Best Local Similarity 66.7%; Pred. NO. 16;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 VPKLDEVYNAAY 14
DB 1 vpkldaaysvay 12

RESULT 11
AAW76485
ID AAW76485 standard; peptide; 12 AA.
XX
AC AAW76485;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #37.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
DR WPI; 1998-522170/45.
XX
PT Modified recombinant allergens - useful for immuno-therapy of
PT allergies
XX
PS Example 1; Page 12; 31pp; German.
XX
CC AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
SQ Sequence 12 AA;

Query Match 15.9%; Score 44; DB 19; Length 12;
Best Local Similarity 66.7%; Pred. NO. 16;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 VPKLDEVYNAAY 14
DB 1 vpkldaaysvay 12

RESULT 11
AAW76485
ID AAW76485 standard; peptide; 12 AA.
XX
AC AAW76485;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #37.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
PA (MERE) MERCK PATENT GMBH.
XX
PI Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
DR WPI; 1998-522170/45.
XX
PT Modified recombinant allergens - useful for immuno-therapy of
PT allergies

PT allergies
PS Example 1; Page 12; 31pp; German.
XX
CC AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
SQ Sequence 12 AA;

Query Match 15.9%; Score 44; DB 19; Length 12;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 39 IIAGTPEVHAVK 50
:|||||
Db 1 viagalelvhavk 12

RESULT 12
AAW10845
ID AAW10845 standard; peptide; 29 AA.
XX
AC AAW10845;
XX
DT 07-NOV-1997 (first entry)
XX
DE MAb anti-HBsAg binder sequence, FC22, from R26 library.
XX
KW Functional surrogate; analyte; affinity receptor; immunoreactive group;
KW mimic; homogenous immunoassay; detection; diagnostic analyte; Chlamydia;
KW cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;
KW pregnancy; infectious disease; ferritin; myosin light chain; troponin;
KW follicle stimulating hormone; human; growth hormone; immunoglobulin E;
KW prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;
KW antibody; chorionic gonadotropin; luteinising hormone; cytomegalovirus;
KW Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;
KW carcinoembryonic antigen; alpha-fetoprotein; herpes virus; Cal25.
XX
OS Synthetic.
XX
PN WO9641172-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US10498.
XX
PR 07-JUN-1995; 95US-0476375.
XX
PA (CYTO-) CYTOGEN CORP.
XX
PI Carter JM, Lee-Own FV;
XX
DR WPI; 1997-077284/07.
XX
XX Labelled functional surrogate of an analyte - useful as competitor
PT molecule in affinity assays, esp. for detecting large macromolecules
PT such as ferritin
XX
PS Claim 51; Page 95; 156pp; English.
XX
CC This sequence represents a monoclonal anti-hepatitis B antigen binder
CC sequence from the R26 library which may be used in the conjugate of the
CC invention. The novel labelled conjugate comprises at least one label
CC attached to a functional surrogate of an analyte of interest. The
CC surrogate is capable of competing effectively with the analyte for a

CC limiting amount of an affinity receptor for the analyte. The conjugate
CC exhibits an activity that is altered upon interaction with the affinity
CC receptor and this activity can be measured and related to the amount of
CC the analyte present in a sample. Functional surrogates such as this have
CC an immunoreactive group that allows the surrogate to compete effectively
CC and with the analyte for a limiting amount of its affinity receptor.
CC Functional surrogates are able to mimic naturally occurring analytes.
CC They can be labelled for use in standard competitive affinity assays
CC (esp. homogenous immunoassays) for detecting large macromolecules such
CC as polypeptides, polysaccharides, polynucleotides, glycoproteins and
CC lipid-containing macromolecules, as well as small haptens. Typical
CC diagnostic analytes for detection include cardiac or tumour markers,
CC allergens, hormones related to fertility-pregnancy or analytes associated
CC with infectious disease. In particular, the assays are useful for
CC detecting ferritin, follicle stimulating hormone, human growth hormone,
CC immunoglobulin E, prolactin, parathyroid hormone, human placental
CC lactogen, hepatitis antigens or antibodies against them, human
CC chorionic gonadotropin, human luteinising hormone, cytomegalovirus,
CC Chlamydia, Streptococcus a, rubella, toxoplasma, herpes virus, DK-MB,
CC myoglobin, myosin light chain, troponin, carcinoembryonic antigen,
CC alpha-fetoprotein, prostate-specific antigen and Cal25 (a tumour marker).
XX
SQ Sequence 29 AA;

Query Match 15.9%; Score 44; DB 18; Length 29;
Best Local Similarity 38.1%; Pred. No. 49;
Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Oy 5 KLDEVYNAAYNAADHAAPEDK 25
::||| |::||| | |
Db 3 eidpfyhyisaadqggsade 23

RESULT 13
AAW10860
ID AAW10860 standard; peptide; 33 AA.
XX
AC AAW10860;
XX
DT 07-NOV-1997 (first entry)
XX
DE MAb anti-HBsAg binder sequence, FC22, from R26 library.
XX
KW Functional surrogate; analyte; affinity receptor; immunoreactive group;
KW mimic; homogenous immunoassay; detection; diagnostic analyte; Chlamydia;
KW cardiac marker; tumour marker; allergen; hormone; fertility; myoglobin;
KW pregnancy; infectious disease; ferritin; myosin light chain; troponin;
KW follicle stimulating hormone; human; growth hormone; immunoglobulin E;
KW prolactin; parathyroid hormone; placental lactogen; hepatitis antigen;
KW antibody; chorionic gonadotropin; luteinising hormone; cytomegalovirus;
KW Streptococcus; rubella; toxoplasma; DK-MB; prostate-specific antigen;
KW carcinoembryonic antigen; alpha-fetoprotein; herpes virus; Cal25.
XX
OS Synthetic.
XX
PN WO9641172-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96WO-US10498.
XX
PR 07-JUN-1995; 95US-0476375.
XX
PA (CYTO-) CYTOGEN CORP.
XX
PI Carter JM, Lee-Own FV;
XX
DR WPI; 1997-077284/07.
XX
XX Labelled functional surrogate of an analyte - useful as competitor
PT molecule in affinity assays, esp. for detecting large macromolecules

PT such as ferritin
 XX
 PS Claim 51; Page 97; 156pp; English.
 XX
 CC This sequence represents a monoclonal anti-hepatitis B antigen binder
 CC sequence from the R26 library which may be used in the conjugate of the
 CC invention. The novel labelled conjugate comprises at least one label
 CC attached to a functional surrogate of an analyte of interest. The
 CC surrogate is capable of competing effectively with the analyte for a
 CC limiting amount of an affinity receptor for the analyte. The conjugate
 CC exhibits an activity that is altered upon interaction with the affinity
 CC receptor and this activity can be measured and related to the amount of
 CC the analyte present in a sample. Functional surrogates such as this have
 CC an immunoreactive group that allows the surrogate to compete effectively
 CC and with the analyte for a limiting amount of its affinity receptor.
 CC Functional surrogates are able to mimic naturally occurring analytes.
 CC They can be labelled for use in standard competitive affinity assays
 CC (esp. homogenous immunoassays) for detecting large macromolecules such
 CC as polypeptides, polysaccharides, polynucleotides, glycoproteins and
 CC lipid-containing macromolecules, as well as small haptens. Typical
 CC diagnostic analytes for detection include cardiac or tumour markers,
 CC allergens, hormones related to fertility-pregnancy or analytes associated
 CC with infectious disease. In particular, the assays are useful for
 CC detecting ferritin, follicle stimulating hormone, human growth hormone,
 CC immunoglobulin E, prolactin, parathyroid hormone, human placental
 CC lactogen, hepatitis antigens or antibodies against them, human
 CC chorionic gonadotropin, human luteinising hormone, cytomegalovirus,
 CC Chlamydia, Streptococcus A, rubella, toxoplasma, herpes virus, PK-MB,
 CC myoglobin, myosin light chain, troponin, carcinoembryonic antigen,
 CC alpha-fetoprotein, prostate-specific antigen and CA125 (a tumour marker).
 XX
 SQ Sequence 33 AA;

Query Match 15.9%; Score 44; DB 18; Length 33;
 Best Local Similarity 38.1%; Pred. No. 58;
 Matches 8; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 5 KLDVYNAAYNAADHAAPEDK 25
 : : | : | : | : | : | : | :
 Db 7 eidpfyhiysaagdsarsde 27

RESULT 14
 AAR80961
 ID AAR80961 standard; peptide; 36 AA.
 AC AAR80961;
 XX
 DT 23-APR-1996 (first entry)
 XX
 DE Integrin alpha-6a cytoplasmic domain.
 XX
 KW Integrin; chimeric; extracellular; transmembrane; cytoplasmic;
 KW ligand; activation inhibitor; inflammation; autoimmune disease;
 KW transplantation; thrombus; cancer.
 XX
 OS Homo sapiens!
 XX
 PN WO9525173-A1.
 XX
 PD 21-SEP-1995.
 XX
 PF 06-MAR-1995; 95WO-US02885.
 XX
 PR 14-MAR-1994; 94US-0214770.
 XX
 XX (SCRI) SCRIPPS RES INST.
 XX
 PI Ginsberg MH, O'Toole TE;
 XX
 DR WPI; 1995-336977/43.
 XX

PT Chimeric integrin mol. comprising reporter and target integrin
 PT domains - for identifying integrin activation inhibitors which are
 PT useful for treating or preventing unwanted immune responses
 XX
 PS Disclosure; Figure 2; 50pp; English.
 XX
 CC Chimeric integrin molecules comprising the extracellular and
 CC transmembrane domains of a reporter integrin (RI) fused to the cyto-
 CC plasmic domain of a target integrin (TI). The RI is pref. from
 CC alphaIIb-beta3 and the TI is pref. chosen from alphaV-beta3, alpha-
 CC beta2, alphaI-beta2, alpha2-beta1, alpha5-beta1, alpha6a-beta1,
 CC alphaIIb-beta3 or alpha4-beta1. This sequence is the cytoplasmic
 CC domain of alpha-2. The chimeric integrins are useful for
 CC identifying TI inhibitors which can be used to treat mammalian
 CC cancers, thrombosis or any unwanted immune response, e.g.
 CC inflammation; autoimmune disea
 XX
 SQ Sequence 36 AA;

Query Match 15.7%; Score 43.5; DB 16; Length 36;
 Best Local Similarity 52.9%; Pred. No. 76;
 Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 10 YNAAYNAAD-HAAPEDK 25
 | : | : | : | : | : | : | : | :
 Db 13 ydatynkaelhaqpsdk 29

RESULT 15
 ABB28520
 ID ABB28520 standard; Peptide; 53 AA.
 XX
 AC ABB28520;
 XX
 DT 01-FEB-2002 (first entry)
 XX
 DE Peptide #1171 encoded by breast cell single exon nucleic acid probe.
 XX
 KW Human; microarray; single exon probe; gene expression; breast;
 KW disease; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157271-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00662.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-496933/54.
 XX
 XX New spatially-addressable set of single exon nucleic acid probes,
 PT useful for measuring gene expression in sample derived from human
 PT breast, comprises number of single exon nucleic acid probes -
 XX
 PS Claim 27; SEQ ID NO 11488; 327pp + sequence listing; English.
 XX
 CC The invention relates to a spatially-addressable set of single exon
 CC nucleic acid probes for measuring gene expression in a sample derived
 CC from human breast and Bt 474 cells. The method involves contacting

CC the probes with a collection of detectably labelled nucleic acids
 CC derived from mRNA of human breast, and then measuring the label
 CC bound to each probe of the microarray. The probes are useful for
 CC verifying the expression of regions of genomic DNA predicted to
 CC encode proteins. They are useful for gene discovery, and for
 CC determining predisposition and/or prognosing breast disease. Gene
 CC expression analysis is useful for assessing the toxicity of chemical
 CC agents on cells. The microarray of this invention presents a far greater
 CC diversity of probes for measuring gene expression, with far less bias
 CC than expressed sequence tag microarrays. The method is suitable for
 CC rapid production of functional information from genomic sequence. The
 CC present sequence is a peptide encoded by a single exon nucleic acid
 CC probe of the invention.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

xx
 SQ Sequence 53 AA;

Query Match 15.5%; Score 43; DB 22; Length 53;
 Best Local Similarity 54.5%; Pred. No. 1.4e+02;
 Matches 12; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 7 DEVYNAAY-NRADHAAPEDKYE 27
 ||| | :||| || |
 Db 33 devgevvyfsaadh-epeskve 53

Search completed: July 11, 2002, 11:00:14
 Job time: 211 sec

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:58:53 ; Search time 12.95 seconds
(without alignments)
99.966 Million cell updates/sec

Title: 'US-09-696-169A-20

Perfect score: 1277

Sequence: 1 QLVKLDVYNAAYNAADHA.....SEALHIAGTPEVHAVKPCA 53

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 165294

Minimum DB seq length: 0

Maximum DB seq length: 53

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*

- 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	65	23.5	20	1	US-08-440-861-12
2	57	20.6	20	1	US-08-440-861-11
3	56	20.2	20	1	US-08-440-861-13
4	54	19.5	20	1	US-08-440-861-14
5	48	17.3	13	1	US-08-433-854-28
6	48	17.3	13	1	US-08-174-745A-28
7	48	17.3	13	2	US-08-195-947-28
8	48	17.3	13	2	US-08-433-885-28
9	48	17.3	13	2	US-08-433-908B-28
10	48	17.3	13	4	US-08-410-614-28
11	48	17.3	20	1	US-08-440-861-15
12	43.5	15.7	36	1	US-08-214-770-8
13	43.5	15.7	36	5	PCT-US95-02885-8
14	41	14.8	20	1	US-08-440-861-8
15	40	14.4	53	4	US-08-900-574-4
16	39.5	14.3	43	1	US-08-178-477B-21
17	39	14.1	20	1	US-08-440-861-7
18	38.5	13.9	35	2	US-08-749-526-6
19	38	13.7	20	3	US-08-817-926-3
20	38	13.7	32	4	US-09-082-279B-1226
21	38	13.7	32	4	US-09-315-304B-1226
22	37	13.4	35	4	US-09-082-279B-784
23	37	13.4	35	4	US-09-315-304B-784
24	37	13.4	45	4	US-08-963-851-10
25	37	13.4	47	4	US-08-936-165A-365
26	37	13.4	48	1	US-08-446-692-37
27	37	13.4	48	2	US-08-488-351A-37

Sequence 9, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 25, Appl
Sequence 3, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 15, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 57, Appl

ALIGNMENTS

RESULT 1
US-08-440-861-12
; Sequence 12, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106.016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-12

Query Match 23.5%; Score 65; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.019;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 10 YNAAYNAADHAAPEDKYEAF 29
|: || ||: | || ||: ||
Db 1 YDVAKRAEGATPEAKYDAF 20

RESULT 2

US-08-440-861-11
; Sequence 11, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-11

Query Match 20.6%; Score 57; DB 1; Length 20;
Best Local Similarity 58.8%; Pred. No. 0.22;
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 2 LVPKLDVYNAAYNAAD 18
|: || ||: | || ||:
Db 3 LIPKLDYDVAKYAAE 19

RESULT 3

US-08-440-861-13
; Sequence 13, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-13

Query Match 20.2%; Score 56; DB 1; Length 20;
Best Local Similarity 55.0%; Pred. No. 0.3;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 20 AAPEDKYEAFVLFHFSALHI 39
| || ||: || ||: || ||:
Db 1 ATPEAKYDAEVTALTEALRV 20

RESULT 4

US-08-440-861-14
; Sequence 14, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:


```
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragoutas
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-14

Query Match 19.5%; Score 54; DB 1; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.56;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 30 VLFSEALHIIAGTPEVHAV 49
| :||| :||| :|||
Db 1 VTALTEALRVIAALEVHAV 20

RESULT 5
US-08-433-854-28
; Sequence 28, Application US/08433854
; Patent No. 5721119
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-440-861-14

Query Match 17.3%; Score 48; DB 1; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
| | | | | | | | | |
Db 1 IAGALEVHAVKP 12

RESULT 6
US-08-174-745A-28
; Sequence 28, Application US/08174745A
; Patent No. 5736362
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5736362ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/174,745A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-174-745A-28

Query Match 17.3%; Score 48; DB 1; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
| | | | | | | | | |
Db 1 IAGALEVHAVKP 12
```

```
RESULT 7
US-08-193-947-28
; Sequence 28, Application US/08195947
; Patent No. 5840316
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avtioglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/195,947
; FILING DATE: 14-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-193-947-28

Query Match 17.3%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
   ||| |||||
Db 1 IAGALEVHAVKP 12

RESULT 8
US-08-433-885-28
; Sequence 28, Application US/08433885
; Patent No. 5869333
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avtioglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
```

```
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,885
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-885-28

Query Match 17.3%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
   ||| |||||
Db 1 IAGALEVHAVKP 12

RESULT 9
US-08-433-908B-28
; Sequence 28, Application US/08433908B
; Patent No. 5965455
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avtioglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,908B
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 436
```

ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-039C2D4
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-908B-28

Query Match 17.3%; Score 48; DB 2; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
Db 1 IAGALEVHAVKP 12

RESULT 10
US-08-410-614-28
Sequence 28, Application US/08410614
Patent No. 6277383
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avjoglu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphioglu, Cenk
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/410,614
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/195,947
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
US-08-410-614-28
Query Match 17.3%; Score 48; DB 4; Length 13;
Best Local Similarity 83.3%; Pred. No. 2.1;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
Db 1 IAGALEVHAVKP 12

RESULT 11
US-08-440-861-15
Sequence 15, Application US/08440861
Patent No. 5710126
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Luqman, Mohammad
TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
TITLE OF INVENTION: ALLERGEN
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,861
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragoras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-15

Query Match 17.3%; Score 48; DB 1; Length 20;
Best Local Similarity 83.3%; Pred. No. 3.5;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 40 IAGTPEVHAVKP 51
Db 1 IAGALEVHAVKP 12

RESULT 12
US-08-214-770-8
Sequence 8, Application US/08214770
Patent No. 5523209

GENERAL INFORMATION:
APPLICANT: Ginsberg, Mark H.
FILING DATE: O'Toole, Tim
TITLE OF INVENTION: METHODS FOR IDENTIFYING
TITLE OF INVENTION: INHIBITORS OF INTEGRIN
TITLE OF INVENTION: ACTIVATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/214,770
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06410/002001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 36
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-214-770-8

Query Match 15.7%; Score 43.5; DB 1; Length 36;
Best Local Similarity 52.9%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 10 YNAAYNAAD-HAAPEDK 25
I:|:|:|:|:|:|
Db 13 YDATYHKAETHAQPSDK 29

RESULT 13
PCT-US95-02885-8
Sequence 8, Application PC/TUS9502885
GENERAL INFORMATION:
APPLICANT: Ginsberg, Mark H.
FILING DATE: O'Toole, Timothy
TITLE OF INVENTION: METHODS FOR IDENTIFYING INHIBITORS
TITLE OF INVENTION: OF INTEGRIN ACTIVATION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: MS-DOS (Version 5.0)
SOFTWARE: WordPerfect (Version 5.1)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/02885
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/214,770
FILING DATE: March 14, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 06410/002001
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 36
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US95-02885-8

Query Match 15.7%; Score 43.5; DB 5; Length 36;
Best Local Similarity 52.9%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 4; Indels 1; Gaps 1;

QY 10 YNAAYNAAD-HAAPEDK 25
I:|:|:|:|:|:|
Db 13 YDATYHKAETHAQPSDK 29

RESULT 14
US-08-440-861-8
Sequence 8, Application US/08440861
Patent No. 5710126
GENERAL INFORMATION:
APPLICANT: Griffith, Irwin J.
APPLICANT: Kuo, Mei-Chang
APPLICANT: Lugman, Mohammad
TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
TITLE OF INVENTION: ALLERGEN
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/440,861
FILING DATE: 15-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,016
FILING DATE: 31-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-440-861-8

Query Match 14.8%; Score 41; DB 1; Length 20;
Best Local Similarity 44.4%; Pred. No. 30;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 19 HAAPEDKYAEFLHFSEA 36
Db 1 NAPPADKKEFEAFSES 18

RESULT 15
US-08-900-574-4
Sequence 4, Application US/08900574
Patent No. 6221837
GENERAL INFORMATION:
APPLICANT: Ertl, Johann
APPLICANT: Habermann, Paul
APPLICANT: Geisen, Karl
APPLICANT: Seipke, Gerhard
TITLE OF INVENTION: Insulin derivatives with increased zinc
TITLE OF INVENTION: binding
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,
ADDRESSEE: & Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: District of Columbia
COUNTRY: U.S.A.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/900,574
FILING DATE: July 24, 1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: German Application No. 6221837 19630242.0
FILING DATE: July 26, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Carol P. Einaudi
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1499-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 53 amino acids
TYPE: Amino acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: Protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..53
US-08-900-574-4

Query Match 14.4%; Score 40; DB 4; Length 53;
Best Local Similarity 36.4%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 4; Indels 6; Gaps 1;

QY 32 HFSEALHIAG-----TPEVH 47
DB 10 HLVEALYLVCGERGFFYTPKTH 31

Search completed: July 11, 2002, 11:00:58
Job time: 125 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:03:13 ; Search time 18.29 Seconds
(without alignments)
299.458 Million cell updates/sec

Title: US-09-696-169A-21

Perfect score: 275

Sequence: 1 MAHKFMVAMFLAVAVVLGL.....LIEDVNASFRAAMATTANVP 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 14670

Minimum DB seq length: 0

Maximum DB seq length: 57

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	80	29.1	28	2 A60359	pollen allergen DG
2	45	16.4	44	2 AC3544	LSU ribosomal prot
3	41.5	15.1	51	2 B45246	orf 3'of ldh - Lac
4	41	14.9	55	2 A45959	endoglucanase B (E
5	40.5	14.7	30	2 T07505	photosystem I prot
6	40.5	14.7	49	2 H75292	hypothetical prote
7	40	14.5	48	2 T20751	hypothetical prote
8	40	14.5	52	2 S63981	immune-inducible p
9	39.5	14.4	47	2 S03775	hypothetical prote
10	39.5	14.4	56	2 A11895	hypothetical prote
11	39	14.2	57	2 G87230	probable secreted
12	38.5	14.0	44	2 S54144	tmsl protein - mou
13	38	13.8	38	2 S61280	nikkomycin synthet
14	38	13.8	44	2 F87344	ribosomal protein
15	38	13.8	55	2 A28911	male accessory gla
16	37.5	13.6	50	2 T06541	probable NADPH-fe
17	37	13.5	33	2 S03945	hydrogen dehydroge
18	37	13.5	45	2 P00599	tyrosine 3-monooxy
19	37	13.5	45	2 P00590	tyrosine 3-monooxy
20	37	13.5	45	2 P00591	tyrosine 3-monooxy
21	37	13.5	45	2 P00592	tyrosine 3-monooxy
22	37	13.5	45	2 P00593	tyrosine 3-monooxy
23	37	13.5	47	1 J00732	ribosomal protein
24	37	13.5	49	2 C71316	hypothetical prote
25	37	13.5	50	1 LBBCB	light-harvesting p
26	37	13.5	55	2 D37334	D-mandelate dehydr
27	36.5	13.3	39	2 S10315	photosystem II pro
28	36.5	13.3	45	2 T52279	R2R3-MVB transcrip
29	36	13.1	33	2 S69595	gene 4f-rnp protei

30 36 13.1 38 2 S76107 photosystem II pro
31 36 13.1 38 2 AB1966 photosystem II pro
32 36 13.1 46 2 C83437 hypothetical prote
33 36 13.1 50 2 D90706 hypothetical prote
34 36 13.1 53 2 H84193 hypothetical prote
35 36 13.1 53 2 B84236 hypothetical prote
36 36 13.1 54 1 BGSB spermatid transiti
37 36 13.1 55 2 A70070 hypothetical prote
38 36 13.1 56 2 E83079 hypothetical prote
39 35.5 12.9 57 2 AH2906 conserved hypothet
40 35 12.7 34 2 S32408 dnaK-type molecula
41 35 12.7 42 2 A95019 hypothetical prote
42 35 12.7 45 2 T52125 R2R3-MVB transcrip
43 35 12.7 47 2 G87679 hypothetical prote
44 35 12.7 51 2 E97394 hypothetical prote
45 35 12.7 54 2 S31918 myrosinase - chick

ALIGNMENTS

RESULT 1

A60359 pollen allergen DG3 - orchard grass (fragment)
C:Species: Dactylis glomerata (orchard grass)
C:Date: 03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change 03-Nov-2000
C:Accession: A60359
R:Walsh, D.J.; Matthews, J.A.; Denmeade, R.; Maxwell, P.; Davidson, M.; Walker, M.R.
Int. Arch. Allergy Appl. Immunol. 91, 419-425, 1990
A:Title: Monoclonal antibodies to proteins from cocksfoot grass (Dactylis glomerata)
A:Reference number: A60359; MUID:91007970
A:Accession: A60359
A:Molecule type: protein
A:Residues: 1-28 <WAL>
C:Superfamily: grass pollen allergen IX
C:Keywords: pollen

Query Match 29.1%; Score 80; DB 2; Length 28;
Best Local Similarity 75.0%; Pred. No. 0.0046;
Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 26 AEGGKATTEQKLIEDVNAS 45

| | | | | | | | | | | | | | | | | | | |

Db 9 AGGKATTEQKLMEDINXA 28

RESULT 2

AC3544 LSU ribosomal protein l34P [Imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AC3544
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AC3544
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-44 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53518.1; PID:gl7984424; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0276
A:Map position: II

Query Match 16.4%; Score 45; DB 2; Length 44;
Best Local Similarity 41.7%; Pred. No. 1.2e+02;
Matches 10; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

Query Match 14.5%; Score 40; DB 2; Length 48;
Best Local Similarity 48.1%; Pred. No. 5.3e+02;
Matches 13; Conservative 6; Mismatches 4; Indels 3;
Gaps 3;

Qy 17 VLGLATSPTEAGGK-ATTEEQ--KLIE 40
||| ||| ||| : ||| ||| |||
Db 23 ILGFA-SPSAQGTESASTRSQAKLIVE 48

RESULT 8
S63981
Immune-inducible proline-rich protein precursor - fruit fly (Drosophila melanogaster)
N:Alternate names: metchnikowin
C:Species: Drosophila melanogaster
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000
C:Accession: S63981; S63980
R:Levashina, E.A.; Ohresser, S.; Bulet, P.; Reichhart, J.M.; Hetru, C.; Hoffmann, J.A.
Eur. J. Biochem. 233, 694-700, 1995
A:Title: Metchnikowin, a novel immune-inducible proline-rich peptide from Drosophila wit
A:Reference number: S63980; MUID:96067716
A:Accession: S63981
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-52 <LEV>
A:Cross-references: GB:X91060; NID:g1052588; PIDN:CAA62511.1; PID:g1052589
A:Note: 19-Thr, 29-Arg was also found
A:Accession: S63980
A:Molecule type: protein
A:Residues: 27-52 <LEW>
A:Note: 3-Arg was also found
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-52/Product: immune-inducible proline-rich protein #status predicted <MAT>

Query Match 14.5%; Score 40; DB 2; Length 52;
Best Local Similarity 52.6%; Pred. No. 5.7e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 9 AMFLAVAVVLGATSPTE 27
||| ||| ||| ||| |||
Db 7 AIFLALLGVMTATSVLAE 25

RESULT 9
S03775
Hypothetical protein (rrnD 5' region) - Streptomyces coelicolor (fragment)
C:Species: Streptomyces coelicolor
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 18-Jun-1993
C:Accession: S03775
R:Baylis, H.A.; Bibb, M.J.
Mol. Microbiol. 2; 569-579, 1988
A:Title: Transcriptional analysis of the 16S rRNA gene of the rrnD gene set of Streptomy
A:Reference number: S03775; MUID:89039245
A:Accession: S03775
A:Molecule type: DNA
A:Residues: 1-47 <BAY>
A:Cross-references: EMBL:Y00411

Query Match 14.4%; Score 39.5; DB 2; Length 47;
Best Local Similarity 34.1%; Pred. No. 5.9e+02;
Matches 15; Conservative 5; Mismatches 13; Indels 11; Gaps 2;

Qy 11 FLAVAVVLGATSPTEAGGKATTEQKLIEDVNASFRAMATTA 54
||| ||| ||| : ||| ||| |||
Db 14 FLAVYVL-----GGRAVDGE--TGDVEAERSATVPTA 46

RESULT 10
A11895
Hypothetical protein as10715 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: A11895
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A11895
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-56 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072672.1; PID:g17130060; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asl0715

Query Match 14.4%; Score 39.5; DB 2; Length 56;
Best Local Similarity 32.1%; Pred. No. 7.1e+02;
Matches 18; Conservative 5; Mismatches 20; Indels 13; Gaps 3;

Qy 6 FMVAMFLAVAVVLGLA----TSPTAEGGKATTEQKLIEDVNASFRAMATTA 57
||| ||| ||| : ||| ||| |||
Db 6 FGFALMLSLAFLGACGETQTPPPAATGEPT-----DTGAATPAA-TTPAATP 52

RESULT 11
G87230
Probable secreted protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87230
R:Coie, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A:Title: Massive gene decay in the leprosy bacillus
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: G87230
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-57 <STO>
A:Cross-references: GB:AL450380; NID:g13093813; PIDN:CAC32101.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2569A

Query Match 14.2%; Score 39; DB 2; Length 57;
Best Local Similarity 29.4%; Pred. No. 8.3e+02;
Matches 10; Conservative 4; Mismatches 14; Indels 6; Gaps 1;

Qy 2 AAHKFVYAMFLAVAVVLGLA-----TSPTAEGG 29
||| ||| ||| : ||| ||| |||
Db 8 AAASVVVLLGLGAATFGMTLMVQDTPKPLPGG 41

RESULT 12
S54144
tms1 protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 08-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-1995
C:Accession: S54144
R:Kustikova, O.S.
submitted to the EMBL Data Library, April 1995
A:Reference number: S54144
A:Accession: S54144
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-44 <KUS>
A:Cross-references: EMBL:X86374

Query Match 14.0%; Score 38.5; DB 2; Length 44;
Best Local Similarity 45.5%; Pred. NO. 7.3e+02;
Matches 10; Conservative 2; Mismatches 5; Indels 5; Gaps 1;

QY 13 AVAVVLGLATS-----PTAEGG 29
I: :||||| | | |
DB 6 ALLALLGLATSCSFICPAVSGG 27

RESULT 13

S61280

nikkomycin synthesis protein P6 - Streptomyces tendae (fragments)

C:Species: Streptomyces tendae

C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 07-May-1999

C:Accession: S61280

R:Moehrlie, V.; Roos, U.; Bormann, C.

Mol. Microbiol. 15, 561-571, 1995

A:Title: Identification of cellular proteins involved in nikkomycin production in Streptomyces tendae

A:Reference number: S61278; MUID:95302967

A:Accession: S61280

A:Molecule type: protein

A:Residues: 1-20;21-38 <MOE>

Query Match 13.8%; Score 38; DB 2; Length 38;
Best Local Similarity 50.0%; Pred. NO. 7.2e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 20 LATSPTAEGGKATTEE 35
| | | | | | | | | |
DB 22 LATPTADVAKIATVED 37

RESULT 14

F87344

Ribosomal protein L34 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: F87344

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A:Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87344

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-44 <STO>

A:Cross-references: GB:AE005673; NID:gl3422006; PIDN:AAK22754.1; GSPDB:GN00148

C:Genetics:

A:Gene: CC0769

Query Match 13.8%; Score 38; DB 2; Length 44;
Best Local Similarity 39.1%; Pred. NO. 8.4e+02;
Matches 9; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 30 KATTEEQKLIEDVNASPRAAMAT 52
| | : | | : | | | |
DB 2 KRTFQPSKLVRRRHGYRARMAT 24

RESULT 15

A28911

male accessory gland peptide precursor - fruit fly (Drosophila melanogaster)

N:Alternate names: paragonial peptide

C:Species: Drosophila melanogaster

C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 05-Dec-1998

C:Accession: A28911

R:Chen, P.S.; Stumm-Zollinger, E.; Aigaki, T.; Balmer, J.; Bienz, M.; Boehlen, P.

Cell 54, 291-299, 1988

A:Title: A male accessory gland peptide that regulates reproductive behavior of female D

A:Reference number: A28911; MUID:88282537

A:Accession: A28911

A:Molecule type: mRNA

A:Residues: 1-55 <CHE>

A:Cross-references: GB:M21201; NID:gl58124; PID:gl58125

A>Note: part of this sequence, including the amino end of the mature protein, was determined by Edman degradation.

C:Genetics:

A:Gene: FlyBase:Acp70A

A:Cross-references: FlyBase:FBgn0003034

C:Keywords: hydroxyproline

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-55/Product: male accessory gland peptide #status experimental <MAT>

F:28,32,34,36,38/Modified site: hydroxyproline (Pro) #status experimental

F:33/Modified site: isoleucine derivative (Ile) #status experimental

F:43-55/Disulfide bonds: #status predicted

Query Match 13.8%; Score 38; DB 2; Length 55;
Best Local Similarity 50.0%; Pred. NO. 1.1e+03;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 8 VAMFLAVAVVLGLATS 23
:|:|:|:| | | | |
DB 4 LALFLVLCVVLGLVQS 19

Search completed: July 11, 2002, 11:05:21
Job time: 128 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:04:58 ; Search time 10.15 Seconds
(without alignments) 217.440 Million cell u

Title: US-09-696-169A-21

perfect score: 275

Sequence: 1 MAAHKFMVAMFLAVAVLGL.....LIEDVNASFRAAMATTANVP 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs. 38719550 residues

Total number of hits satisfying chosen parameters: 4574

Minimum DB seq length: 0

Minimum DB	seq	length:
Maximum DB	seq	length: 5'

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	40.5	14.7	30	1	PSAM_PINTH
2	40.5	14.7	44	1	PH61_CHRSP
3	40.5	14.7	44	1	YZ04_BACSU
4	40	14.5	52	1	MTK_DROME
5	39	14.2	57	1	Y23A_MYCLE
6	38	13.8	44	1	RL34_CAUCR
7	38	13.8	55	1	A70A_DROME
8	37	13.5	33	1	HOXU_RHOOP
9	37	13.5	46	1	CSPA_PHOMO
10	37	13.5	47	1	RL34_PROMI
11	37	13.5	49	1	LHB_ERYSIP
12	37	13.5	49	1	Y495_TREPA
13	37	13.5	57	1	Y23A_MVCTU
14	36.5	13.3	38	1	PSB1_SYNP6
15	36	13.1	39	1	PSB1_SYNP3
16	36	13.1	54	1	STP1_SHEEP
17	35	12.7	34	1	HS7S_CUCMA
18	35	12.7	47	1	YQ54_BACAN
19	35	12.7	56	1	Y546_METJA
20	34.5	12.5	42	1	RK34_OLILU
21	34.5	12.5	50	1	HOKC_ECOLI
22	33.5	12.2	31	1	PEPL_PORPU
23	33.5	12.2	31	1	PSAM_CHILVU
24	33.5	12.2	48	1	TRPD_BACME
25	33.5	12.2	45	1	LHB1_RHOGE
26	33.5	12.2	50	1	VG38_BPMD2
27	33	12.0	33	1	Y474_BORLU
28	33	12.0	45	1	RL34_CHLMU
29	33	12.0	48	1	YC21_ARCFU
30	33	12.0	56	1	SCP2_MESMA
31	33	12.0	56	1	SCP3_WESMA
32	32.5	11.8	55	1	COP6_STAAU
33	32.5	11.8	55	1	LHB_RHOTE

ALIGNMENTS

```

RESULT 1
PSAM_PINTH STANDARD; PRT; 30 AA.
ID PSAM_PINTH STANDARD; PRT; 30 AA.
AC P41601;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Photosystem I reaction centre subunit XII (Psi-M).
PSAM.
OS Pinus thunbergii (Green pine) (Japanese black pine).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus.
NCBI_TaxID=3350;
[1]
SEQUENCE FROM N.A.
RP MEDLINE=95024047; PubMed=7937893;
RX Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T.,
RA Wakisuga M.;
RT "Loss of all ndh genes as determined by sequencing the entire
RT chloroplast genome of the black pine Pinus thunbergii." ;
RL Proc. Natl. Acad. Sci. U.S.A. 91:9794-9798(1994).
CC -1- SIMILARITY: BELONGS TO THE PSAM FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; D17510; BAA04383.1; -
DR EMBL; D17510; BAA04316.1; -
DR Photosystem I; Photosynthesis; Chloroplast.
SQ SEQUENCE 30 AA; 3321 MW; A150441AB5D5AF7C CRC64;

```

Query Match 14.7%; Score 40.5; DB 1; Length 30;
Best Local Similarity 42.3%; Pred. No. 1.3e+02;
Matches 11; Conservative 4; Mismatches 6; Indels 5; Gaps 1;

Qy 1 MAAHKENVAMFLA-----VAVVLGLA 21
| |||:| ||| :|: |||
Db 1 MVELKELIAFFLAFTAGILAIKLGOA 26

RESULT	2				
PHEI_CHRSP					
ID	PHEI_CHRSP	STANDARD;	PRT;	44	AA.
AC	P30942;				
DT	01-JUL-1993	(Rel. 26, Created)			
DT	01-JUL-1993	(Rel. 26, Last sequence update)			
DT	15-JUL-1998	(Rel. 36, Last annotation update)			
DE	Phycocerythrin alpha-1 chain (Fragment).				
OS	Chromoceras sp.				

OC Eukaryota; Cryptophyta; Cryptomonadaceae; Chromonas.
OX NCBI_TaxID=3029;

RN [1]
RP SEQUENCE.
RC STRAIN=CS 24;
RX MEDLINE=91032178; PubMed=2226853;
RA Jenkins J., Hiller R.G., Speirs J., Godovac-Zimmermann J.:
RT "A genomic clone encoding a cryptophyte phycoerythrin alpha-subunit.
RT Evidence for three alpha-subunits and an N-terminal membrane transit
RT sequence.";
RL FEBS Lett. 273:191-194(1990).
CC -1- FUNCTION: LIGHT-HARVESTING PHOTOSYNTHETIC BILE PIGMENT-PROTEIN
CC FROM THE PHCOBLIPROTEIN COMPLEX.
CC -1- SUBUNIT: THE ORIGIN OF MULTIPLE FORMS OF PHYCOERYTHRIN MAY LIE IN
CC THE ALPHA SUBUNITS. THE 2(1ALPHA/BETA) STRUCTURE COULD COMPRISE ANY
CC COMBINATION OF 2 OUT OF 4 DIFFERENT ALPHA UNITS WITH AN INVARIANT
CC BETA UNIT.
CC -1- SUBCELLULAR LOCATION: WITHIN THE THYLAKOID LUMEN. PERIPHERY OF THE
CC RODS OF THE PHCOBLISOME.
CC -1- PTM: CONTAINS ONE COVALENTLY LINKED BILIN CHROMOPHORE.
CC InterPro: IPR004228; phycoerythr_ab.
DR Pfam: PF02972; phycoerythr_ab.1.
KW Phycobilisome; Electron transport; Photosynthesis; Bile pigment;
KW Chloroplast.
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4723 MW; 2EBE2D2EAF47AA7E CRC64;

Query Match 14.7%; Score 40.5; DB 1; Length 44;
Best Local Similarity 34.8%; Pred. No. 2e+02;
Matches 8; Conservative 8; Mismatches 6; Indels 1; Gaps 1;

Oy 19 GLATSPTAE-GCKATPERQKLE 40
| : : | : | | : : | : :
Db 18 GCSRAPKSETGTATKDDOMVK 40

RESULT 3
YZQA_BACSU
ID YZQA_BACSU STANDARD; PRT; 44 AA.
AC P40771;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein clone PSP28 (Fragment).
GN YZQA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89108019; PubMed=3145906;
RA Smith H., de Jong A., Bron S., Venema G.:
RT "Characterization of signal-sequence-coding regions selected from the
RT Bacillus subtilis chromosome.";
RL Gene 70:351-361(1986).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M22915; AAA22831.1; -
DR Subtilist; BG11031; yzqa.
KW Hypothetical protein; Complete proteome.
FT NON_TER 44 44
SQ SEQUENCE 44 AA; 4696 MW; 548FA6DF26755CA1 CRC64;

Query Match 14.7%; Score 40.5; DB 1; Length 44;
Best Local Similarity 39.4%; Pred. No. 2e+02;
Matches 13; Conservative 5; Mismatches 12; Indels 3; Gaps 1;

Oy 5 KFWAMFLAVAVVGLATSP---ABGGRATTE 34
| | | | | | | | : : | : :
Db 4 KKKVMCLAVTLVFGMSPTLTNSGGFKESTD 36

RESULT 4
MTK_DROME
ID MTK_DROME STANDARD; PRT; 52 AA.
AC Q24395; Q24396; Q9V7B9;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Metchnikowin precursor.
GN MTK OR CG8175.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 27-52.
RC STRAIN=OREGON-R; TISSUE=Abdomen, and Thorax;
RX MEDLINE=96067716; PubMed=7588819;
RA Levashina E.A., Ohresser S., Bulet B., Reichhart J.-M., Hetru C.,
RA Hoffmann J.A.;
RT "Metchnikowin, a novel immune-inducible proline-rich peptide from
RT Drosophila with antibacterial and antifungal properties.";
RL Eur. J. Biochem. 233:694-700(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=98263241; PubMed=9600835;
RA Levashina E.A., Ohresser S., Lemaitre B., Imler J.-L.;
RT "Two distinct pathways can control expression of the gene encoding
RT the Drosophila antimicrobial peptide metchnikowin.";
RL J. Mol. Biol. 278:515-527(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Glaser K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Flosser C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Thewissen C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.:
RA "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: POTENT ANTIFUNGAL AND ANTIBACTERIAL ACTIVITY AGAINST
CC GRAM-POSITIVE BACTERIA.
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN FAT BODY. MAYBE BLOOD
CC CELLS PARTICIPATE IN THE PRODUCTION OF THE PEPTIDE.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED RAPIDLY AND STRONGLY AT ALL STAGES.
CC -!- POLYMORPHISM: 2 ALLELIC FORMS (A1 AND A2) VARYING IN TWO AA
CC POSITION. THE ISOFORM SHOWN HERE IS A1
CC -!- SIMILARITY: TO DIPTERICIN, HEMIPERICIN, DROSOCIN, APIDABICINS AND
CC TO THE C-TERMINUS OF ABAECIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X91060; CAA62511.1; -
DR EMBL; X91061; CAA62512.1; -
DR EMBL; AF030959; AAC64659.1; -
DR EMBL; AE003811; AAF58139.1; -
DR FlyBase; FBgn0014865; Mtk.
KW Insect immunity; Antibiotic; Fungicide; Signal.
FT SIGNAL 1 24
FT PROPEP 25 26
FT PEPTIDE 27 52
FT VARIANT 19 19
FT VARIANT 29 29
FT SEQUENCE 52 AA; 5654 MW; 0B1E2112BAE03129 CRC64;
SQ
Query Match 14.5%; Score 40; DB 1; Length 52;
Best Local Similarity 52.6%; Pred. No. 2.7e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 9 AMFLAVAVVLGLATSPTAEE 27
DB 7 AIFLALLGVMTATSVLAE 25
RESULT 5
ID Y23A_M1CYCLE STANDARD; PRT; 57 AA.
AC Q9CD20;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative secreted protein ML2569.1 precursor.
GN ML2569.1 OR ML2569A.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
```

```
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RA "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL583926; CAC32101.1; -
DR Leproma; ML2569A; -
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 32
FT CHAIN 33 57
FT SEQUENCE 57 AA; 5897 MW; 7CC82EB212AC94C1 CRC64;
SQ
Query Match 14.2%; Score 39; DB 1; Length 57;
Best Local Similarity 29.4%; Pred. No. 3.8e+02;
Matches 10; Conservative 4; Mismatches 14; Indels 6; Gaps 1;
QY 2 AAHKFWVMFLAVAVVLGLA-----TSPTAEGG 29
DB 8 AAASVVVGLLGAATIFGTMVLVQDTPPLPGG 41
RESULT 6
ID RL34_CAUCR STANDARD; PRT; 44 AA.
AC P58129;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L34.
GN RPMH OR CC0769.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group.
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- SIMILARITY: BELONGS TO THE L34P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005752; AAK22754.1; -
DR TIGR; CC0769; -
DR InterPro; IPR000271; Ribosomal_L34.
DR Pfam; PF00468; Ribosomal_L34; 1.
DR PROSITE; P500784; RIBOSOMAL_L34; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
```

[illegible]


```
RESULT 11
LHB_ERYSP
AC P26274; STANDARD; PRT; 49 AA.
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Light-harvesting protein B-870, beta chain (Antenna pigment protein,
DE beta chain).
GN PUPB.
OS Erythrobacter sp. (strain OCH 114).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Roseobacter.
OX NCBI_TaxID=2434;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92157872; PubMed=1787796;
RA Liebetanz R., Hornberger U., Drews G.;
RT "Organization of the genes coding for the reaction-centre L and M
RT subunits and B870 antenna polypeptides alpha and beta from the
RT aerobic photosynthetic bacterium Erythrobacter species OCH114.";
RL Mol. Microbiol. 5:1459-1468(1991).
CC -!- FUNCTION: ANTENNA COMPLEXES ARE LIGHT-HARVESTING SYSTEMS, WHICH
CC TRANSFER THE EXCITATION ENERGY TO THE REACTION CENTERS.
CC -!- SUBUNIT: THE CORE COMPLEX IS FORMED BY DIFFERENT ALPHA AND BETA
CC CHAINS, BINDING BACTERIOCHLOROPHYLL MOLECULES, AND ARRANGED
CC MOST PROBABLY IN TETRAMERIC STRUCTURES DISPOSED AROUND THE
CC REACTION CENTER. THE NON-PIGMENTED GAMMA CHAINS MAY CONSTITUTE
CC ADDITIONAL COMPONENTS.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57597; CAA40815.1; -.
DR PIR; S16309; LBCB.
DR HSSP; P02951; IDX7.
DR InterPro; IPR002362; Antenna_comp_beta.
DR InterPro; IPR000066; LHC.
DR Pfam; PF00556; LHC; 1.
DR PRINTS; PR00674; LIGHTHARVSTB.
DR PRODOM; PD001979; Antenna_comp_beta; 1.
DR PROSITE; PS00969; ANTENNA_COMP_BETA; 1.
KW Antenna complex; Light-harvesting polypeptide; Transmembrane;
KW Magnesium; Bacteriochlorophyll; Inner membrane.
FT INIT_MET 0
FT BY_SIMILARITY 0
FT DOMAIN 1 21 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 22 44 POTENTIAL.
FT DOMAIN 45 49 PERIPLASMIC (POTENTIAL).
FT BINDING 20 20 BACTERIOCHLOROPHYLL (POTENTIAL).
FT METAL 38 38 AXIAL LIGAND TO THE BACTERIOCHLOROPHYLL
FT MAGNESIUM (POTENTIAL).
SQ SEQUENCE 49 AA; 5461 MW; 929B906BE4559369 CRC64;

Query Match 13.5%; Score 37; DB 1; Length 49;
Best Local Similarity 69.2%; Pred. No. 5.7e+02;
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 10 MFLAVAVLGLAT 22
:| | | | | | | |
Db 29 LFAAVAVVAHLAT 41

RESULT 12
Y495_TREPA
ID Y495_TREPA STANDARD; PRT; 49 AA.
AC O83508;
```

```
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0495.
GN TP0495.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF001226; AAC65490.1; -.
DR TIGR; TP0495; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 49 AA; 5166 MW; 01E3617D5E5DB39C CRC64;

Query Match 13.5%; Score 37; DB 1; Length 49;
Best Local Similarity 34.0%; Pred. No. 5.7e+02;
Matches 16; Conservative 4; Mismatches 13; Indels 14; Gaps 2;

QY 7 MVAMFLAV-----AVVLGLATPTAEGG-----KATTEOKLI 39
:| | | | | | | | | | | | | | | | | |
Db 1 MLMLRLAVPCRONAVSYLLALPLPKGIFSVALAVSGTAESRLV 47

RESULT 13
Y23A_MYCTU
ID Y23A_MYCTU STANDARD; PRT; 57 AA.
AC P58240;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative secreted protein RV0236.1 precursor.
GN RV0236.1 OR RV0236A OR MT0250.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
```


Search completed: July 11, 2002, 11:08:23
Job time: 205 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:04:38 ; Search time 24.17 Seconds
(without alignments)
407.973 Million cell updates/sec

Title: US-09-696-169A-21

Perfect score: 275

Sequence: 1 MAHKFWAMFLAVAVVLG.....LIEDVNASFRAMATTANVP 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 37469

Minimum DB seq length: 0

Maximum DB seq length: 57

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP_TREMBL.19.*
2: sp_archaea.*
3: sp_bacteria.*
4: sp_fungi.*
5: sp_human.*
6: sp_invertebrate.*
7: sp_mammal.*
8: sp_mhc.*
9: sp_organelle.*
10: sp_phase.*
11: sp_plant.*
12: sp_rodent.*
13: sp_virus.*
14: sp_vertebrate.*
15: sp_unclassified.*
16: sp_rvirus.*
17: sp_bacteriap.*
18: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	44.5	16.2	44	2 P71476	P71476 lactobacill
2	44	16.0	44	2 Q45266	Q45266 bradyrhizob
3	42.5	15.5	41	8 Q95826	Q95826 galaxias sp
4	42	15.3	43	10 Q95808	Q95808 zea mays (m
5	42	15.3	44	16 Q98090	Q98090 rhizobium l
6	41.5	15.1	52	16 Q92007	Q92007 rhizobium m
7	41	14.9	39	10 Q9AYH2	Q9AYH2 oryza sativ
8	41	14.9	46	6 O19186	O19186 ovilis aries
9	41	14.9	49	6 Q9TV94	Q9TV94 bos taurus
10	41	14.9	53	15 Q87297	Q87297 chimpanzee
11	41	14.9	55	2 Q46042	Q46042 cellulomona
12	40.5	14.7	49	16 Q9RS51	Q9RS51 deinococcus
13	40	14.5	37	10 Q92P82	Q92P82 raphanus sa
14	40	14.5	38	2 Q9F5C7	Q9F5C7 agrobacteri
15	40	14.5	45	10 Q49801	Q49801 arabidopsis
16	40	14.5	48	5 O17790	O17790 caenorhabdi

17	39	14.2	41	13 Q9PSL4	Q9PSL4 torpedo cal
18	39	14.2	44	16 Q92SF3	Q92SF3 rhizobium m
19	38.5	14.0	47	2 Q34026	Q34026 burkholderi
20	38	13.8	37	10 Q9ZP81	Q9ZP81 raphanus sa
21	38	13.8	42	10 Q9XIT3	Q9XIT3 glycine max
22	38	13.8	45	10 Q49763	Q49763 arabidopsis
23	38	13.8	45	10 Q49781	Q49781 arabidopsis
24	38	13.8	45	10 Q49785	Q49785 arabidopsis
25	38	13.8	45	10 Q9SAM4	Q9SAM4 arabidopsis
26	38	13.8	50	2 Q9ZIL5	Q9ZIL5 streptococc
27	38	13.8	53	15 Q87300	Q87300 chimpanzee
28	38	13.8	56	2 Q50345	Q50345 mycoplasma
29	37.5	13.6	49	12 Q9VYJ5	Q9VYJ5 mucosal dis
30	37	13.5	30	11 Q9JHF4	Q9JHF4 rattus norv
31	37	13.5	32	12 Q9IBM0	Q9IBM0 rabbit hemo
32	37	13.5	41	13 Q9PSL5	Q9PSL5 torpedo cal
33	37	13.5	41	13 Q9PSL3	Q9PSL3 torpedo cal
34	37	13.5	43	4 Q9UF20	Q9UF20 homo sapien
35	37	13.5	43	10 Q9S7U6	Q9S7U6 zea mays (m
36	37	13.5	43	10 Q9SQE0	Q9SQE0 zea mays (m
37	37	13.5	43	10 Q9SQD3	Q9SQD3 zea mays (m
38	37	13.5	48	5 Q9TX32	Q9TX32 strongyloce
39	37	13.5	53	15 Q87306	Q87306 chimpanzee
40	37	13.5	53	15 Q87405	Q87405 chimpanzee
41	37	13.5	53	15 Q87411	Q87411 chimpanzee
42	37	13.5	53	15 Q87414	Q87414 chimpanzee
43	37	13.5	55	2 Q9R554	Q9R554 acinetobact
44	37	13.5	56	15 Q87417	Q87417 chimpanzee
45	37	13.5	57	13 Q9Y6K6	Q9Y6K6 gadus morhu

ALIGNMENTS

RESULT 1
ID P71476 PRELIMINARY; PRT; 44 AA.
AC P71476;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE PLNV (FRAGMENT).
GN PLNV.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C11;
RX MEDLINE=94161498; PubMed=8117074;
RA Diep D.B., Haavarstein S.L., Nissen-Meyer J., Nes I.F.;
RT "The gene encoding plantaricin A, a bacteriocin from lactobacillus
RT plantarum C11, is located on the same transcription unit as an agr-
RT like regulatory system."
RL Appl. Environ. Microbiol. 60:160-166(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=C11;
RX MEDLINE=96414465; PubMed=8817486;
RA Diep D.B., Haavarstein S.L., Nes I.F.;
RT "A bacteriocin-like peptide induces bacteriocin synthesis in
RT Lactobacillus plantarum C11."
RN Mol. Microbiol. 18:631-639(1995).
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=C11;
RX MEDLINE=96345611; PubMed=8755874;
RA Diep D.B., Haavarstein S.L., Nes I.F.;
RT "Characterization of the locus responsible for the bacteriocin
RT production in Lactobacillus plantarum C11."
RL J. Bacteriol. 178:4472-4483(1996).
EMBL; X94434; CAA64194.1; -.


```

RESULT 9
Q9TV94 ID Q9TV94 PRELIMINARY; PRT; 49 AA.
AC Q9TV94;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
DE CALPASTATIN (FRAGMENT).
GN CAST.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Palmer B.R., Su H.-Y., Roberts N., Hickford J.G.H., Bickerstaffe R.;
RT "Single Nucleotide Polymorphisms in an Intron of the Ovine Calpastatin
Gene.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF117813; AAD21054.1; -.
FT NON_TER 1
FT NON_TER 49
SQ SEQUENCE 49 AA; 4992 MW; 54941AC8BFA32362 CRC64;

Query Match 14.9%; Score 41; DB 6; Length 49;
Best Local Similarity 33.3%; Pred. No. 8.3e+02;
Matches 12; Conservative 8; Mismatches 14; Indels 2; Gaps 1;

QY 22 TSPTAEGGKATTEOKLIEDVNASFRAAMATTANVP 57
:||||: | |||: | : | : |
Db 16 SSPTADAKK--TEKEKSTEEALKQAGSAGVIRSAAPP 49

RESULT 10
Q87297 ID Q87297 PRELIMINARY; PRT; 53 AA.
AC Q87297;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ENVELOPE GLYCOPROTEIN (FRAGMENT).
GN ENV.
OS Chimpanzee immunodeficiency virus (SIV(cpz)) (CIV).
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11723;
RN [1]
RP SEQUENCE FROM N.A.
RA Chackerian B., Morton W.R., Overbaugh J.;
RT "Persistence of simian immunodeficiency virus Mne variants upon
transmission.";
RL J. Virol. 68:4080-4085(1994).
DR EMBL; u06278; AAA20249.1; -.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
KW AIDS; Coat protein; Glycoprotein.
FT NON_TER 1
FT NON_TER 53
SQ SEQUENCE 53 AA; 5587 MW; 003E55865FDEFF27 CRC64;

Query Match 14.9%; Score 41; DB 15; Length 53;
Best Local Similarity 40.7%; Pred. No. 9.1e+02;
Matches 11; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

QY 19 GLATSPTEAGGKATTEOKLIEDVNAS 45
||| | | | | | | | | | | | | |
Db 14 GLTKSSTTTAPTATSTEAIAIKVNES 40

```

```

RESULT 11
Q46042 ID Q46042 PRELIMINARY; PRT; 55 AA.
AC Q46042;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE ENDOGLUCANASE B (CENB) PRECURSOR (FRAGMENT).
OS Cellulomonas fimi.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Cellulomonadaceae; Cellulomonas.
OX NCBI_TaxID=1708;
RN [1]
RP SEQUENCE FROM N.A.
RA Owolabi J.B., Beguin P., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
RT "Expression in Escherichia coli of the Cellulomonas fimi structural
gene for endoglucanase B.";
RL Appl. Environ. Microbiol. 54:518-523(1988).
DR EMBL; M33026; AAA23085.1; -.
DR HSSP; P26221; ITF4.
DR InterPro; IPR001701; Glyco_hydro_9.
DR Pfam; PF00759; Glyco_hydro_9; 1.
KW Signal.
FT SIGNAL 1 33 POTENTIAL.
FT CHAIN 34 >55 POTENTIAL.
FT NON_TER 55
SQ SEQUENCE 55 AA; 5608 MW; 37B3CCD04A72393D CRC64;

Query Match 14.9%; Score 41; DB 2; Length 55;
Best Local Similarity 40.0%; Pred. No. 9.5e+02;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 7 MVAMFLAVAVVLGLATSPTEAGGKA 31
:| | | | | | | | | | | | | | |
Db 9 LVAGGSALAVAVGVLVPLATGAA 33

RESULT 12
Q9RS51 ID Q9RS51 PRELIMINARY; PRT; 49 AA.
AC Q9RS51;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 5.1 KDA PROTEIN.
GN DR2276.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Sheu M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002060; AAF11832.1; -.
DR TIGR; DR2276; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 49 AA; 5071 MW; 83A28B018D2EB501 CRC64;

Query Match 14.7%; Score 40.5; DB 16; Length 49;
Best Local Similarity 58.8%; Pred. No. 9.6e+02;

```

```
Matches 10; Conservative 3; Mismatches 3; Indels 1; Gaps 1;
QY 10 MELAVAVVLGLATSPTA 26
    :|||:|:|:|:|
Db 16 IFLVVALVWGM-TGPTA 31

RESULT 13
Q92P82
ID Q92P82 PRELIMINARY; PRT; 37 AA.
AC Q92P82
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE COLD SHOCK PROTEIN (FRAGMENT).
GN CSP5.
OS Raphanus sativus (Radish).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Raphanus.
OX NCBI_TaxID=3726;
RN [1]
RA Baymiev A.K., Gimalov F.R., Vakhitov V.A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132903; CAB39172.1;
FT NON_TER 1
FT NON_TER 37
SQ SEQUENCE 37 AA; 3575 MW; D7A14C4606C28311 CRC64;

Query Match 14.5%; Score 40; DB 10; Length 37;
Best Local Similarity 34.5%; Pred. No. 8e+02;
Matches 10; Conservative 2; Mismatches 17; Indels 0; Gaps 0;
QY 25 TAEGGKATTEOKLIEDVNASFRAAMATT 53
    |||:|:|:|:|:|:|
Db 8 TAAGASATAGOKISEAAVNVVKEKT 36

RESULT 14
Q9F5C7
ID Q9F5C7 PRELIMINARY; PRT; 38 AA.
AC Q9F5C7
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE RIORF125 PROTEIN.
GN RIORF125.
OS Agrobacterium rhizogenes.
OG Plasmid prl1724.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=359;
RN [1]
RA "The complete nucleotide sequence of a Ri (root inducing) plasmid
RA indicates its chimerical structure between Ti and Sym plasmids."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Maeda Y., Satou M., Satuti N., Kataoka M., Tanaka N.,
RA Yoshida K.;
RT "Analysis of unique variable region of a plant root inducing plasmid,
RT prl1724, by the construction of its physical map and library."
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RA Moriguchi K., Nishida T., Maeda Y., Tanaka N., Yoshida K.;
```

```
RT "Genome structure of Ri plasmid (1): Construction of linking library
RT and physical map of prl1724 in Japanese Agrobacterium.";
RL Nucleic Acids Symp. Ser. 39:189-190(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF03-01724;
RX MEDLINE=20241294; PubMed=10780382;
RA Maeda Y., Moriguchi K., Kataoka M., Satou M., Satuti N., Tanaka N.,
RA Yoshida K.;
RT "Genome structure of Ri plasmid (1): Sequencing analysis of T-DNA and
RT its flanking regions of prl1724 in Japanese Agrobacterium
RT rhizogenes.";
RL Nucleic Acids Symp. Ser. 42:67-68(1999).
DR EMBL: AP002086; BAB16244.1;
KW Plasmid.
SQ SEQUENCE 38 AA; 3711 MW; E9B74EBFB795D0A3 CRC64;

Query Match 14.5%; Score 40; DB 2; Length 38;
Best Local Similarity 39.4%; Pred. No. 8.3e+02;
Matches 13; Conservative 7; Mismatches 9; Indels 4; Gaps 2;
QY 7 MVAMFLAVAVVLGLAT---SPTAEGGK-ATTEE 35
    :||:|:|:|:|:|:|
Db 6 LIALVLALAGASAAATVLIIVKSSDGGKPAITEE 38

RESULT 15
O49801
ID O49801 PRELIMINARY; PRT; 45 AA.
AC O49801
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE R2R3-MYB TRANSCRIPTION FACTOR (FRAGMENT).
GN ATMVB79.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RA "One hundred R2R3-MYB genes in the genome of Arabidopsis thaliana."
RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z95802; CAB09234.1;
DR HSSP: Q03237; 1A5J.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam: PF00249; myb_DNA-binding; 1.
DR PROSITE: PS50090; MYB_3; 1.
FT NON_TER 1
FT NON_TER 45
SQ SEQUENCE 45 AA; 5351 MW; C4F16570A5772049 CRC64;

Query Match 14.5%; Score 40; DB 10; Length 45;
Best Local Similarity 29.0%; Pred. No. 1e+03;
Matches 9; Conservative 8; Mismatches 14; Indels 0; Gaps 0;
QY 24 PTAEGGKATTEOKLIEDVNASFRAAMATTA 54
    |:|:|:|:|:|:|:|:|
Db 9 PDLKRGQITPHEESIILDLHAKGNRWSTIA 39

Search completed: July 11, 2002, 11:08:07
Job time: 209 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:01:03 ; Search time 29.64 Seconds
(without alignments)
213.603 Million cell updates/sec

Title: US-09-696-169A-21

Perfect score: 275

Sequence: 1 MAHKFWAMFLAVAVVLGL.....LIEDVNASFRAAMATTANVP 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 375566

Minimum DB seq length: 0

Maximum DB seq length: 57

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	82	29.8	20	16 AAR71510	LPX-3, peptide fr
2	71	25.8	20	16 AAR71511	LPX-4, peptide fr
3	65	23.6	20	17 AAR89385	Kentucky Blue Gras
4	58	21.1	12	19 AAR76455	Graminae pollen al
5	53	19.3	12	19 AAR76456	Kentucky pollen al
6	53	19.3	20	17 AAR89386	Kentucky Blue Gras
7	51	18.5	20	15 AAR62917	TCR alpha chain pe
8	49	17.8	12	19 AAR76454	Graminae pollen al
9	48	17.5	20	16 AAR71509	LPX-2.1, peptide
10	48	17.5	27	20 AAY09286	Rice beta-glucanas
11	47	17.1	56	21 AAB33392	Eucalyptus grandis

12	46	16.7	12	19 AAR76457	Graminae pollen al
13	45	16.4	33	22 AAO03546	Human polypeptide
14	45	16.4	57	22 AAM85676	Human immune/haema
15	43	15.6	24	17 AAR707320	C-terminus of myc-
16	43	15.6	24	20 AAY12699	Human 5' EST seque
17	43	15.6	44	5 AAP40042	Sequence encoded b
18	43	15.6	50	21 AAY86235	Human secreted pro
19	42	15.3	15	22 AAB86663	Human cytomegalovi
20	42	15.3	27	22 AAM18617	Peptide #5051 enco
21	42	15.3	27	22 AAM31077	Peptide #5114 enco
22	42	15.3	39	11 AAR07270	Recombinant-derive
23	42	15.3	52	21 AAB38350	Human secreted pro
24	42	15.3	53	18 AAW19369	Beta 7 integrin R1
25	42	15.3	53	20 AAY33887	Beta 7 integrin su
26	42	15.3	56	22 AAU16602	Human novel secret
27	41	14.9	12	19 AAR76458	Graminae pollen al
28	41	14.9	35	22 AAG77424	Human colon cancer
29	41	14.9	36	16 AAR82536	IA beta chain frag
30	41	14.9	41	21 AAG47131	Arabidopsis thalia
31	40.5	14.7	20	17 AAR89396	Kentucky Blue Gras
32	40.5	14.7	45	8 AAR70040	Secretory signal s
33	40.5	14.7	56	22 AAU67017	Propionibacterium
34	40.5	14.7	56	22 AAU54696	Propionibacterium
35	40	14.5	36	22 ABB29709	Peptide #2360 enco
36	40	14.5	36	22 ABB34881	Peptide #2387 enco
37	40	14.5	36	22 ABB20298	Protein #2297 enco
38	40	14.5	36	22 AAM55691	Human brain expres
39	40	14.5	36	22 AAM15893	Peptide #2327 enco
40	40	14.5	36	22 AAM28401	Peptide #2438 enco
41	40	14.5	36	22 AAM03628	Peptide #2310 enco
42	40	14.5	43	16 AAR82539	IA beta chain frag
43	40	14.5	45	22 AAG19054	Novel human diagno
44	40	14.5	51	21 AAG22556	Zea mays proteain f
45	40	14.5	52	21 AAY91545	Human secreted pro

ALIGNMENTS

RESULT 1

AAR71510
ID AAR71510 standard; Protein; 20 AA.
XX
AC AAR71510;
XX
DT 02-NOV-1995 (first entry)
XX
DE LPX-3, peptide fragment of Lol pv protein allergen.
XX
DE Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
KW sensitivity; ryegrass pollen allergen.
KW Lolium perenne.
OS
XX
PN W09506728-A.
XX
PD 09-MAR-1995.
XX
PF 05-AUG-1994; 94WO-US09024.
XX
PR 13-AUG-1993; 93US-0106016.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Griffith IJ, Kuo M, Luqman M;
XX
DR WPI; 1995-115444/15.
XX
PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
PT and DNA - for treating sensitivity to rye-grass pollen allergen
PT or an immunologically cross-reactive allergen.
XX
PS Claim 1; Fig 2; 110pp; English.

```

XX CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
CC T cell epitope derived from the Lol pv protein are claimed, and can be
CC used to treat or diagnose sensitivity to ryegrass pollen in an
CC individual or to pollen proteins that are immunologically related to Lol
CC pv, such as Dac gv (see AAR71507).
XX SQ Sequence 20 AA;

Query Match 29.8%; Score 82; DB 16; Length 20;
Best Local Similarity 84.2%; Pred. No. 0.00017;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 26 AEGGKATTEOKLIEDVNA 44
   1 |||||:||||:||||
Db 2 aaggtatdeqklliedvna 20

RESULT 2
AAR71511
ID AAR71511 standard; Protein; 20 AA.
XX AC AAR71511;
XX DT 02-NOV-1995 (first entry)
XX DE LPX-4, peptide fragment of Lol pv protein allergen.
XX KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
XX KW sensitivity; ryegrass pollen allergen.
XX OS Lolium perenne.
XX PN W09506728-A.
XX PD 09-MAR-1995.
XX PF 05-AUG-1994; 94WO-US09024.
XX PR 13-AUG-1993; 93US-0106016.
XX PA (IMMU-) IMMULOGIC PHARM CORP.
XX PI Griffith IJ, Kuo M, Luqman M;
XX DR WPI; 1995-115444/15.
XX PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
XX PT and DNA - for treating sensitivity to rye-grass pollen allergen
XX PT or an immunologically cross-reactive allergen.
XX PS Claim 1; Fig 2; 110pp; English.
XX CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
CC T cell epitope derived from the Lol pv protein are claimed, and can be
CC used to treat or diagnose sensitivity to ryegrass pollen in an
CC individual or to pollen proteins that are immunologically related to Lol
CC pv, such as Dac gv (see AAR71507).
XX SQ Sequence 20 AA;

Query Match 25.8%; Score 71; DB 16; Length 20;
Best Local Similarity 70.0%; Pred. No. 0.0058;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 35 EOKLIEDVNASFRAAMATTA 54
   1|||:|||||:|||||

```

```

Db 1 eqklliedvnaagfkaavaaaa 20

RESULT 3
AAR89385
ID AAR89385 standard; peptide; 20 AA.
XX AC AAR89385;
XX DT 17-SEP-1996 (first entry)
XX DE Kentucky Blue Grass pollen allergen rKBG60 residues 29-48.
XX KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
XX KW allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
XX KW allergic; allergy; prevention; development; reaction; grass;
XX KW antibody; anergise; producing cell; diagnosis; human; clone 60;
XX KW immunological response; B-cell; hlgG; hlgG; T-cell.
XX OS Poa pratensis.
XX PN W09603106-A2.
XX PD 08-FEB-1996.
XX PF 26-JUL-1995; 95WO-CA00439.
XX PR 26-JUL-1994; 94US-0280455.
XX PA (UYMA-) UNIV MANITOBA.
XX PI Mohapatra SS;
XX XX WPI; 1996-116753/12.
XX DR
XX PT Peptide including at least 1 human antigenic determinant of Kentucky
XX PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,
XX PT used to desensitise allergic individuals, prevent grass pollen
XX PT allergy and deplete allergen-specific antibodies
XX PS Claim 6; Page 29; 59pp; English.
XX CC The present peptide is a human antigenic determinant contg. peptide,
XX CC comprising residues 29-48 of the recombinant Kentucky Blue Grass
XX CC pollen allergen rKBG60, which is from the Poa p IX gp. of grass
XX CC pollen allergens. A compsn. comprising the peptide and a carrier
XX CC can be used to desensitise allergic individuals, and prevent
XX CC allergic individuals from developing an allergic reaction to grass
XX CC pollen. The compsn. may also be used to deplete allergen specific
XX CC antibodies from an individual, and anergise allergen specific
XX CC antibody producing cells. The peptide can be used to diagnose grass
XX CC pollen allergy, and has the following human immunological response
XX CC profile, B-cell (hlgG (+)/hlgE (-)), T-cell (+).
XX SQ Sequence 20 AA;

Query Match 23.6%; Score 65; DB 17; Length 20;
Best Local Similarity 66.7%; Pred. No. 0.039;
Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 30 KATTEOKLIEDVNASFR 47
   ||||:|||||:|:|
Db 3 kattdeqklliekinvgfk 20

RESULT 4
AAR76455
ID AAR76455 standard; peptide; 12 AA.
XX AC AAR76455;
XX DT 11-DEC-1998 (first entry)

```


CC The present peptide is a human antigenic determinant contg. peptide,
 CC comprising residues 39-58 of the recombinant Kentucky Blue Grass
 CC pollen allergen KRG60, which is from the Poa p IX gp. of grass
 CC pollen allergens. A compsn. comprising the peptide and a carrier
 CC can be used to desensitise allergic individuals, and prevent
 CC allergic individuals from developing an allergic reaction to grass
 CC pollen. The compsn. may also be used to deplete allergen specific
 CC antibodies from an individual, and anergise allergen specific
 CC antibody producing cells. The peptide can be used to diagnose grass
 CC pollen allergy, and has the following human immunological response
 CC profile, B-cell (hlgG (-)/hlgE (-)), T-cell (-).
 XX
 SQ Sequence 20 AA;

Query Match 19.3%; Score 53; DB 17; Length 20;

Best Local Similarity 50.0%; Pred. No. 1.8;
 Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 38 LIEDVNASFRAAMATTANVP 57
 ||| : | : || : | |
 Db 1 liekinvgfkaavaagvvp 20

RESULT 7

AAR62917
 ID AAR62917 standard; Protein; 20 AA.

XX AC AAR62917;

XX DT 12-JUN-1995 (first entry)

XX DE TCR alpha chain peptide control.

XX KW Allergen; immunotherapy; T-cell receptor; TCR; CDR;

XX KW complementarity determining region; immunosuppressive;

XX KW antibody engineering; suppressor T-lymphocyte; Ts; ovalbumin;

XX KW monoclonal antibody; myeloma; IgG.

XX OS Synthetic.

XX PN W09425489-A.

XX PD 10-NOV-1994.

XX PF 26-APR-1994; 94WO-CA00228.

XX PR 26-APR-1993; 93GB-0008581.

XX FA (UYMA-) UNIV MANITOBA.

XX PI Mohapatra SS, Sehon AH;

XX DR WPI; 1994-358193/44.

XX PT Synthetic peptide(s) for effecting immuno:therapy - have an amino
 PT acid sequence comprising a portion of the CDR3 region of a T-cell
 PT receptor

XX PS Disclosure; Page 44; 78pp; English.

XX CC Antibody production against allergenic antigen (Ag) is specifically
 CC suppressed by treatment with the junctional segment of the alpha
 CC and/or beta chain of the Ag receptor of suppressor T (Ts) cells
 CC (TCR) induced by tolerogenic Ag-PEG conjugates. Induction of
 CC antigen-specific suppression by treatment of mice with TCR
 CC alpha-chain peptides was demonstrated using the peptides given in
 CC AAR62912-17. The peptide of AAR62917 was used as a control.

XX SQ Sequence 20 AA;

Query Match 18.5%; Score 51; DB 15; Length 20;

Best Local Similarity 45.0%; Pred. No. 3.4;
 Matches 9; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 38 LIEDVNASFRAAMATTANVP 57
 : || : | : || : | |
 Db 1 miekinvgfkaavaagvvp 20

RESULT 8

AAW76454
 ID AAW76454 standard; peptide; 12 AA.

XX AC AAW76454;

XX DT 11-DEC-1998 (first entry)

XX DE Graminae pollen allergen Phl p 5b peptide fragment #6.

XX KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 KW epitope; immunotherapy; allergy; hyposensitisation.

XX OS Graminae.

XX PN DE19713001-A1.

XX PD 01-OCT-1998.

XX PF 27-MAR-1997; 97DE-1013001.

XX PR 27-MAR-1997; 97DE-1013001.

XX PA (MERE) MERCK PATENT GMBH.

XX PI Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;

XX PI Kahlert H, Mueller W, Schramm G, Stuewe H;

XX DR WPI; 1998-522170/45.

XX PT Modified recombinant allergens - useful for immuno-therapy of
 PT allergies

XX PS Example 1; Page 12; 31pp; German.

XX CC AAW76449-W76534 are peptide fragments of a natural pollen allergen,
 CC Phl p 5b, isolated from Graminae species. This allergen can be modified
 CC and the reactivity of the modified allergens with IgE antibodies to
 CC grass pollen allergens is reduced or eliminated while their reactivity
 CC with T cells is retained. The genes for the allergens are modified so
 CC that the encoded polypeptides have one or more amino acid substitutions,
 CC deletions and/or additions. The dominant T-cell epitopes of the
 CC allergens are not genetically altered. Such allergens have applications
 CC in the immunotherapy of allergies e.g. hyposensitisation.

XX SQ Sequence 12 AA;

Query Match 17.8%; Score 49; DB 19; Length 12;

Best Local Similarity 100.0%; Pred. No. 3.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 29 GRATTEEQKL 38
 |||||
 Db 3 gkatteeqkl 12

RESULT 9

AAR71509
 ID AAR71509 standard; Protein; 20 AA.

XX AC AAR71509;

XX DT 02-NOV-1995 (first entry)

DE LPIX-2.1, peptide fragment of Lol pv protein allergen.
 XX Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 KW sensitivity; ryegrass pollen allergen.
 XX Lolium perenne.

OS WO9506728-A.
 PN XX
 XX
 XX
 PD 09-MAR-1995.

XX 05-AUG-1994; 94WO-US09024.
 XX 13-AUG-1993; 93US-0106016.
 XX (IMMU-) IMMULOGIC PHARM CORP.

XX Griffith IJ, Kuo M, Luqman M;
 XX WPI; 1995-115444/15.
 XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 PT and DNA - for treating sensitivity to rye-grass pollen allergen
 PT or an immunologically cross-reactive allergen.
 XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).
 XX Sequence 20 AA;

Query Match 17.5%; Score 48; DB 16; Length 20;
 Best Local Similarity 64.3%; Pred. NO. 8.7;
 Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 21 ATSPTEGGKATTE 34
 Db 7 aatpaaaggkattd 20

RESULT 10
 AAY09286
 ID AAY09286 standard; Protein; 27 AA.

XX AAY09286;

XX 07-JUL-1999 (first entry)

XX Rice beta-glucanase gene Gns5 signal peptide.

DE Rice; beta-glucanase; Gns; promoter; plant resistance; expression;
 KW fungal infection; transgenic monocotyledon; growth.

XX Oryza sativa.

OS WO9859046-A1.

XX 30-DEC-1998.

XX 25-JUN-1998; 98WO-US13525.

XX 25-JUN-1997; 97US-0050675.

XX (REGC) UNIV CALIFORNIA.

XX Rodriguez RL;

XX
 DR
 DR
 XX

WPI: 1999-105620/09.
 N-PSDB; AAX33549.
 New polynucleotides which hybridise with rice beta-glucanase genes -
 useful for transforming monocot plants for various characteristics
 including increased resistance to fungal infection

PT
 PT
 XX
 PS Disclosure; Page 39; 90pp; English.

XX The present invention describes isolated polynucleotides (PN) comprising
 CC a sequence which hybridises under high stringency with a rice
 CC beta-glucanase gene. Also described are: (A) a chimeric gene (CG1) used
 CC to produce a transgenic monocot plant, comprising: (i) a transcriptional
 CC regulatory region which hybridises under high stringency with a rice beta
 CC -glucanase promoter; and (ii) a DNA sequence heterologous to the
 CC and (iii) a second DNA sequence encoding a protein to be produced by the plant;
 CC linked 5' to 3', so that the signal polypeptide is in translation frame
 CC with the protein, and is effective to facilitate secretion of the protein
 CC across aleurone or scutellar epithelium layers into the endosperm of
 CC seeds obtained from the plant; (B) a monocot plant stably transformed
 CC with CG1; (C) seeds from the above plant; (D) a method of enhancing the
 CC resistance of a monocot plant to fungal infection by stably transforming
 CC it with CG1; (E) a method of producing a heterologous protein, comprising
 CC stably transforming a plant with CG1, obtaining seeds from the
 CC transformed plant, germinating the seeds and obtaining the protein from
 CC the seed endosperm; (F) an isolated protein having the characteristics of
 CC a rice beta-glucanase enzyme, corresponding to beta-glucanases 2-9
 CC (Gns 2-9). The polynucleotides of the invention are useful to transform
 CC monocot plants. The polynucleotides and the proteins produced may also
 CC provide increased resistance to fungal infection, improved growth
 CC characteristics, and high expression levels of heterologous proteins in
 CC various tissues obtained from the plant.
 XX Sequence 27 AA;

Query Match 17.5%; Score 48; DB 20; Length 27;
 Best Local Similarity 52.0%; Pred. NO. 13;
 Matches 13; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

OY 1 MAAHKFWAMFLAVVVLGATSPT 25
 Db 1 makhg--vasvitlaivlgvaapt 23

RESULT 11
 AAB33392

ID AAB33392 standard; Protein; 56 AA.

XX AAB33392;

XX 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor protein sequence #458.

DE Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bz1p; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KW homeodomain zipper; LIM domain; AP2; EREBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB.

OS Eucalyptus grandis.

XX WO200053724-A2.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000WO-US06112.

XX 11-MAR-1999; 99US-0266513.

PR 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
XX WPT: 2000-579369/54.
DR
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT
XX
XX Claim 8; Page 742; 747pp; English.
PS
XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is one such
CC transcription factor. The transcription factor may be used to produce a
CC plant having modified gene expression such as a woody plant e.g. a
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
CC to modify the activity of a polypeptide in a plant. The transcription
CC factors of the present invention are members from the following families
CC of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
CC zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
CC Cys2His2, CCAAT box elements and MYB.
XX
XX Sequence 56 AA;
SQ

Query Match 17.1%; Score 47; DB 21; Length 56;
Best Local Similarity 43.5%; Pred. No. 46;
Matches 10; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 24 PTAEGGKATTEEQKLIEDVNASF 46
| | | | | | | | : : : :
Db 1 pdvrrgnitteeqllimelhakw 23

RESULT 12
AAW76457
ID AAW76457 standard; peptide; 12 AA.
XX
AC AAW76457;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #9.
XX
XX Pollen; allergen; Phl p 5b; IGE; antibody; grass; reactivity: T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
XX (MERE) MERCK PATENT GMBH.
PA
XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
XX WPI: 1998-522170/45.
DR
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
PT
XX Exemplar 1; Page 12; 31pp; German.
PS

XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IGE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 12 AA;
SQ

Query Match 16.7%; Score 46; DB 19; Length 12;
Best Local Similarity 66.7%; Pred. No. 8.5;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 36 QKLIEDVNASFR 47
| | | | | | | :
Db 1 qkliedinvgfk 12

RESULT 13
AAO03546
ID AAO03546 standard; Protein; 33 AA.
XX
AC AAO03546;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 17438.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI: 2001-514838/56.
DR N-PSDB; AAI83477.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 20; SEQ ID NO 17438; 1399pp + Sequence Listing; English.
PS
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

Oy 24 PTAEGGKATTEOKLIED 41
Db 4 plqpggstssreqklise 21

Search completed: July 11, 2002, 11:04:37
Job time: 214 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 11:02:53 ; Search time 12.87 Seconds
(without alignments)
108.179 Million cell updates/sec

Title: US-09-696-169A-21

Perfect score: 275
Sequence: 1 MAAHFMVAMFLAVAVVLGL.....LIEDVNASFRAAMATTANVP 57

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 166652

Minimum DB seq length: 0
Maximum DB seq length: 57

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	82	29.8	20	1	US-08-440-861-5
2	71	25.8	20	1	US-08-440-861-6
3	67	24.4	34	1	US-08-433-854-7
4	67	24.4	34	1	US-08-174-745A-7
5	67	24.4	34	2	US-08-195-947-7
6	67	24.4	34	2	US-08-433-885-7
7	67	24.4	34	2	US-08-433-908B-7
8	67	24.4	34	4	US-08-410-614-7
9	66	24.0	21	1	US-08-433-854-11
10	66	24.0	21	1	US-08-174-745A-11
11	66	24.0	21	2	US-08-195-947-11
12	66	24.0	21	2	US-08-433-885-11
13	66	24.0	21	2	US-08-433-908B-11
14	66	24.0	21	4	US-08-410-614-11
15	63	22.9	33	1	US-08-440-861-54
16	53	19.3	12	1	US-08-433-854-16
17	53	19.3	12	1	US-08-174-745A-16
18	53	19.3	12	2	US-08-195-947-16
19	53	19.3	12	2	US-08-433-885-16
20	53	19.3	12	2	US-08-433-908B-16
21	53	19.3	12	4	US-08-410-614-16
22	49	17.8	50	3	US-09-053-197A-67
23	49	17.8	50	4	US-09-085-761A-72
24	48	17.5	12	1	US-08-433-854-17
25	48	17.5	12	1	US-08-174-745A-17
26	48	17.5	12	2	US-08-195-947-17
27	48	17.5	12	2	US-08-433-885-17

28 48 17.5 12 2 US-08-433-908B-17 Sequence 17, Appl
29 48 17.5 12 4 US-08-410-614-17 Sequence 17, Appl
30 48 17.5 20 1 US-08-440-861-4 Sequence 4, Appl
31 48 17.5 27 4 US-09-105-390-24 Sequence 24, Appl
32 42 15.3 39 6 5332671-14 Patent No. 5332671
33 42 15.3 41 6 5194596-21 Patent No. 5194596
34 42 15.3 41 6 5219739-26 Patent No. 5219739
35 42 15.3 53 2 US-08-779-113-51 Sequence 51, Appl
36 41 14.9 12 1 US-08-433-854-15 Sequence 15, Appl
37 41 14.9 12 1 US-08-174-745A-15 Sequence 15, Appl
38 41 14.9 12 2 US-08-195-947-15 Sequence 15, Appl
39 41 14.9 12 2 US-08-433-885-15 Sequence 15, Appl
40 41 14.9 12 2 US-08-433-908B-15 Sequence 15, Appl
41 41 14.9 12 4 US-08-410-614-15 Sequence 15, Appl
42 40 14.5 43 5 PCT-US95-02689-50 Sequence 50, Appl
43 40 14.5 55 4 US-08-858-207A-290 Sequence 290, Appl
44 39.5 14.4 38 4 US-09-117-121-16 Sequence 16, Appl
45 39.5 14.4 38 4 US-09-117-121-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-440-861-5
; Sequence 5, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
; US-08-440-861-5

Query Match 29.8%; Score 82; DB 1; Length 20;
Best Local Similarity 84.2%; Pred. No. 3.5e-05;
Matches 16; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 26 AEGGKATTEQKLIEDVNA 44
| | | | | : | | | | : | | | |
Db 2 AAGGKATTDEQKLEEDVNA 20

RESULT 2

US-08-440-861-6
; Sequence 6, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440.861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
US-08-440-861-6

Query Match 25.8%; Score 71; DB 1; Length 20;
Best Local Similarity 70.0%; Pred. No. 0.0012;
Matches 14; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 35 EOKLIEDVNASFRAMATTA 54
| | | | | : | | | | : | | | |
Db 1 EOKLIEDVNAFGKAAVAAAA 20

RESULT 3

US-08-433-854-7
; Sequence 7, Application US/08433854
; Patent No. 5721119
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn

; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,854
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 34 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-854-7

Query Match 24.4%; Score 67; DB 1; Length 34;
Best Local Similarity 72.2%; Pred. No. 0.009;
Matches 13; Conservative, 2; Mismatches 3; Indels 0; Gaps 0;

Qy 21 ATSPTAEGGKATTEQKL 38
| | | | | : | | | | : | | | |
Db 17 AATPAAGGKATTDEQKL 34

RESULT 4

US-08-174-745A-7
; Sequence 7, Application US/08174745A
; Patent No. 5736362
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terryn
; APPLICANT: Suphloglu, Cenk
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5736362ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/174,745A
APPLICATION NUMBER: US/08/174,745A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-174-745A-7

Query Match 24.4%; Score 67; DB 1; Length 34;
Best Local Similarity 72.2%; Pred. No. 0.009;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 21 ATSPATGGKATTEQKL 38
| : | | | | | : | | | | |
Db 17 AATPAAGGKATTEQKL 34

RESULT 5
US-08-195-947-7
Sequence 7, Application US/08195947
Patent No. 5840316

GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avjoglou, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphioglu, Cenk
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316ris
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,947
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-195-947-7

Query Match 24.4%; Score 67; DB 2; Length 34;
Best Local Similarity 72.2%; Pred. No. 0.009;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 21 ATSPATGGKATTEQKL 38
| : | | | | | : | | | | |
Db 17 AATPAAGGKATTEQKL 34

RESULT 6
US-08-433-885-7
Sequence 7, Application US/08433885
Patent No. 5869333
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avjoglou, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphioglu, Cenk
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,885
FILING DATE: 02-MAY-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-039C2D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-885-7

Query Match 24.4%; Score 67; DB 2; Length 34;
Best Local Similarity 72.2%; Pred. No. 0.009;
Matches 13; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 21 ATSPATGGKATTEQKL 38
| : | | | | | : | | | | |

Db 17 AATPAAGGKATTDEQKL 34

RESULT 7

US-08-433-908B-7

; Sequence 7, Application US/08433908B

; Patent No. 5965455

; GENERAL INFORMATION:

; APPLICANT: Singh, Mohan Bir

; APPLICANT: Knox, Robert B.

; APPLICANT: Smith, Penelope

; APPLICANT: Avjoglu, Asil

; APPLICANT: Theerakulpisut, Piyada

; APPLICANT: Hough, Terryn

; APPLICANT: Suphloglu, Cenik

; APPLICANT: Ong, Eng Kok

; TITLE OF INVENTION: Ryegrass Pollen Allergen

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Lahive & Cockfield, LLP

; STREET: 28 State Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/433,908B

; FILING DATE: 02-MAY-1995

; CLASSIFICATION: 436

; ATTORNEY/AGENT INFORMATION:

; NAME: Mandragouras, Amy E.

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: IMI-039C2D4

; TELEPHONE: (617) 227-7400

; TELEFAX: (617) 742-4214

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 34 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-433-908B-7

Query Match 24.4%; Score 67; DB 2; Length 34;

Best Local Similarity 72.2%; Pred. No. 0.009; 2; Mismatches 3; Indels 0;

Matches 13; Conservative 2; Gaps 0;

QY 21 ATSPTEAGGKATTEQKL 38

I : I I I I I I I I I I

Db 17 AATPAAGGKATTDEQKL 34

RESULT 8

US-08-410-614-7

; Sequence 7, Application US/08410614

; Patent No. 6277383

; GENERAL INFORMATION:

; APPLICANT: Singh, Mohan Bir

; APPLICANT: Knox, Robert B.

; APPLICANT: Smith, Penelope

; APPLICANT: Avjoglu, Asil

; APPLICANT: Theerakulpisut, Piyada

; APPLICANT: Hough, Terryn

; APPLICANT: Suphloglu, Cenik

; APPLICANT: Ong, Eng Kok

; TITLE OF INVENTION: Ryegrass Pollen Allergen

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/410,614

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/195,947

; FILING DATE: 14-FEB-1994

; APPLICATION NUMBER: US 07/930,060

; FILING DATE: 14-AUG-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Hohenschutz, Liza D.

; REGISTRATION NUMBER: 33,712

; REFERENCE/DOCKET NUMBER: IMPH-0024

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (215) 568-3100

; TELEFAX: (215) 568-3439

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 34 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; US-08-410-614-7

Query Match 24.4%; Score 67; DB 4; Length 34;

Best Local Similarity 72.2%; Pred. No. 0.009; 2; Mismatches 3; Indels 0;

Matches 13; Conservative 2; Gaps 0;

QY 21 ATSPTEAGGKATTEQKL 38

I : I I I I I I I I I I

Db 17 AATPAAGGKATTDEQKL 34

RESULT 9

US-08-433-854-11

; Sequence 11, Application US/08433854

; Patent No. 5721119

; GENERAL INFORMATION:

; APPLICANT: Singh, Mohan Bir

; APPLICANT: Knox, Robert B.

; APPLICANT: Smith, Penelope

; APPLICANT: Avjoglu, Asil

; APPLICANT: Theerakulpisut, Piyada

; APPLICANT: Hough, Terryn

; APPLICANT: Suphloglu, Cenik

; APPLICANT: Ong, Eng Kok

; TITLE OF INVENTION: Ryegrass Pollen Allergen

; NUMBER OF SEQUENCES: 53

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5721119ris

; STREET: One Liberty Place - 46th Floor

; CITY: Philadelphia

; STATE: PA

; COUNTRY: USA

; ZIP: 19103

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

QY 23 SPTAEGGKATTEQKL 38
:| | | | | | | | | | |
Db 6 TPAAGGKATTEQKL 21

RESULT 12

US-08-433-885-11
; Sequence 11, Application US/08433885
; Patent No. 5869333
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terry
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433.885
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-885-11

Query Match 24.0%; Score 66; DB 2; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 23 SPTAEGGKATTEQKL 38
:| | | | | | | | | | |
Db 6 TPAAGGKATTEQKL 21

RESULT 13

US-08-433-908B-11
; Sequence 11, Application US/08433908B
; Patent No. 5965455
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terry
; APPLICANT: Suphioglu, Cenik

; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lahive & Cockfield, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/433,908B
; FILING DATE: 02-MAY-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Mandragouras, Amy E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IMI-039C2D4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-433-908B-11

Query Match 24.0%; Score 66; DB 2; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 23 SPTAEGGKATTEQKL 38
:| | | | | | | | | | |
Db 6 TPAAGGKATTEQKL 21

RESULT 14

US-08-410-614-11
; Sequence 11, Application US/08410614
; Patent No. 6277383
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglu, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terry
; APPLICANT: Suphioglu, Cenik
; APPLICANT: Ong, Eng Kok
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6277383ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/410,614

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/433,854
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-433-854-11

Query Match 24.0%; Score 66; DB 1; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 23 SPTAEGGKATTEEQKL 38
:| | | | | | | | | |
Db 6 TPAAGGKATTEEQKL 21

RESULT 10
US-08-174-745A-11
Sequence 11; Application US/08174745A
Patent No. 5736362
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avjologlu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphioglu, Cenk
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5736362rls
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/174,745A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100

TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-174-745A-11

Query Match 24.0%; Score 66; DB 1; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 23 SPTAEGGKATTEEQKL 38
:| | | | | | | | | |
Db 6 TPAAGGKATTEEQKL 21

RESULT 11
US-08-195-947-11
Sequence 11; Application US/08195947
Patent No. 5840316
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avjologlu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
APPLICANT: Suphioglu, Cenk
APPLICANT: Ong, Eng Kok
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5840316rls
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,947
FILING DATE: 14-FEB-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,060
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Hohenschutz, Liza D.
REGISTRATION NUMBER: 33,712
REFERENCE/DOCKET NUMBER: IMPH-0024
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-195-947-11

Query Match 24.0%; Score 66; DB 2; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
;
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/195,947
; APPLICATION NUMBER: 14-FEB-1994
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: US 07/930,060
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohenschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0024
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-410-614-11

Query Match 24.0%; Score 66; DB 4; Length 21;
Best Local Similarity 81.2%; Pred. No. 0.0066;
Matches 13; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 23 SPTAEGGKATTEQKL 38
Db 6 TPAAGGKATTEQKL 21

RESULT 15
US-08-440-861-54
; Sequence 54, Application US/08440861
; Patent No. 5710126
; GENERAL INFORMATION:
; APPLICANT: Griffith, Irwin J.
; APPLICANT: Kuo, Mei-Chang
; APPLICANT: Luqman, Mohammad
; TITLE OF INVENTION: T CELL EPITOPES OF RYEGRASS POLLEN
; TITLE OF INVENTION: ALLERGEN
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/440,861
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,016
; FILING DATE: 31-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: IPC-075 (IMI-040cp)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 33 amino acids
; TYPE: amino acid
```

```
;
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: N-terminal
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7, 13, 16, 20,
; OTHER INFORMATION: /label=Pro is 4Hyp
US-08-440-861-54

Query Match 22.9%; Score 63; DB 1; Length 33;
Best Local Similarity 70.6%; Pred. No. 0.031;
Matches 12; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 21 ATSPTAEGGKATTEQK 37
Db 17 AATPAAGGKATTEQK 33

Search completed: July 11, 2002, 11:04:56
Job time: 123 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:52:12 ; Search time 29.77 Seconds
(without alignments)
212.671 Million cell updates/sec

Title: US-09-696-169A-19
Perfect score: 297
Sequence: 1 SKAPQLVPKLDVYNAAYNA.....SEALHIIAGTPEVHAVKPGA 57

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 375566

Minimum DB seq length: 0
Maximum DB seq length: 57

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	297	100.0	57	AA25637	Phleum sp. allerg
2	70	23.6	20	AAR89393	Kentucky Blue Gras
3	65	21.9	20	AAR71517	LP1X-10, peptide f
4	62	20.9	20	AAR71516	LP1X-9, peptide fr
5	56	18.9	20	AAR71518	LP1X-11, peptide f
6	54	18.2	20	AAR71519	LP1X-12, peptide f
7	49	16.5	20	AAR89392	Kentucky Blue Gras
8	48	16.2	20	AAR71520	LP1X-13, peptide f
9	47	15.8	20	AAR89391	Kentucky Blue Gras
10	45	15.2	12	AAW76471	Graminae pollen al
11	44.5	15.0	53	AAG93042	C glutamicum prote

12	44	14.8	12	19	AAW76473	Graminae pollen al
13	44	14.8	12	19	AAW76485	Graminae pollen al
14	44	14.8	20	16	AAR71515	LP1X-8, peptide fr
15	44	14.8	29	18	AAW10845	MAB anti-HBsAg bin
16	44	14.8	33	18	AAW10860	MAB anti-HBsAg bin
17	43.5	14.6	36	16	AAR80961	Integrin alpha-6a
18	43	14.5	53	22	ABR70488	Drosophila melanog
19	43	14.5	53	22	ABR28520	Peptide #1171 enco
20	43	14.5	53	22	ABR33705	Peptide #1211 enco
21	43	14.5	53	22	ABR19154	Protein #1153 enco
22	43	14.5	53	22	AAW54473	Human brain expres
23	43	14.5	53	22	AAW66878	Human bone marrow
24	43	14.5	53	22	AAW14740	Peptide #1174 enco
25	43	14.5	53	22	AAW27163	Peptide #1200 enco
26	43	14.5	53	22	AAW02462	Peptide #1144 enco
27	42.5	14.3	56	20	AAW41313	Human secreted pro
28	42	14.1	12	19	AAW76472	Graminae pollen al
29	42	14.1	57	22	AAW91429	Human immune/haema
30	41	13.8	20	16	AAR71513	LP1X-6, peptide fr
31	41	13.8	27	22	ABG27510	Novel human diagno
32	41	13.8	53	20	AAW42432	Amino acid residue
33	40	13.5	12	19	AAW76486	Graminae pollen al
34	40	13.5	32	19	AAW47366	Insulin 1 B-chain.
35	40	13.5	51	19	AAW79410	Staphylococcus aur
36	40	13.5	51	21	ABR32429	Human secreted pro
37	40	13.5	54	22	ABR67490	Drosophila melanog
38	39	13.1	12	19	AAW76484	Graminae pollen al
39	39	13.1	20	16	AAR71512	LP1X-5, peptide fr
40	39	13.1	20	17	AAR89394	Kentucky Blue Gras
41	39	13.1	34	19	AAW66394	Cationic peptide A
42	39	13.1	34	21	AAV91693	Cationic peptide A
43	39	13.1	35	22	AAW51195	E. coli AMP gene a
44	39	13.1	39	22	AAW99854	Physcomitrella pat
45	39	13.1	40	22	ABR41608	Peptide #9114 enco

ALIGNMENTS

RESULT 1

AAW25637
ID AAY25637 standard; protein; 57 AA.

AC AAY25637;

DT 30-SEP-1999 (first entry)

DE Phleum sp. allergen Phl p 6 protein fragment #2.

XX Major histocompatibility complex; class II; desensitising; human;
KW allergen; grass; tree; weed; pollen; fungi; mould; insect; sting;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

OS Phleum sp.

PN WO9934826-A1.

PD 15-JUL-1999.

PF 11-JAN-1999; 99WO-GB00080.

PR 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

PI Kay AB, Larche M;

XX WPI; 1999-458255/38.

XX

PT Desensitizing patients to polypeptide allergens

PS Example 6; Page 65; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of Tenbrio mollitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a Phleum sp. (Timothy grass) Phl p 6 allergen.

XX Sequence 57 AA;

SQ

Query Match 100.0%; Score 297; DB 20; Length 57;
Best Local Similarity 100.0%; Pred. No. 2.6e-32;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SKAPOLVKLDVYNAAYNAADHAAPEDKYFAVLHFSALHIIAGTPEVHAVKPCA 57
|||||
DB 1 skaplvkldvynaynaadhaapedkyfaivhfsalhliaagtpevhavkpga 57

RESULT 2

AAR89393

ID AAR89393 standard; peptide; 20 AA.

AC AAR89393;

XX

XX 17-SEP-1996 (first entry)

DT

XX

DE Kentucky Blue Grass pollen allergen rKBG60 residues 129-148.

XX

KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
allergic; allergy; prevention; development; reaction; grass;
antibody; anergise; producing cell; diagnosis; human; clone 60;
immunological response; B-cell; hlgG; hlgE; T-cell.

OS Poa pratensis.

XX

XX WO9603106-A2.

XX

PD 08-FEB-1996.

XX

PF 26-JUL-1995; 95WO-CA00439.

XX

PR 26-JUL-1994; 94US-0280455.

XX

PA (UYMA-) UNIV MANITOBA.

XX

XX Mohapatra SS;

PI

DR WPI; 1996-116753/12.

XX

PT Peptide including at least 1 human antigenic determinant of Kentucky Blue Grass allergen - from Poa p IX gp. of grass pollen allergens, used to desensitise allergic individuals, prevent grass pollen allergy and deplete allergen-specific antibodies

XX

PS Example 5; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide.

CC comprising residues 129-148 of the recombinant Kentucky Blue Grass pollen allergen rKBG60, which is from the Poa p IX gp. of grass pollen allergens. A compsn. comprising the peptide and a carrier can be used to desensitise allergic individuals, and prevent allergic individuals from developing an allergic reaction to grass pollen. The compsn. may also be used to deplete allergen specific antibodies from an individual, and anergise allergen specific antibody producing cells. The peptide can be used to diagnose grass pollen allergy, and has the following human immunological response profile, B-cell (hlgG (-)/hlgE (-)), T-cell (-).

XX Sequence 20 AA;

SQ

Query Match 23.6%; Score 70; DB 17; Length 20;
Best Local Similarity 83.3%; Pred. No. 0.016;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 38 SEALHIIAGTPEVHAVKP 55
|||||
DB 2 sealriiagtlevhgvkp 19

RESULT 3

AAR71517

ID AAR71517 standard; Protein; 20 AA.

XX AAR71517;

AC AAR71517;

XX

DT 02-NOV-1995 (first entry)

XX

DE LP1X-10, peptide fragment of Lol pv protein allergen.

XX

KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
sensitivity; ryegrass pollen allergen.

KW

XX Lolium perenne.

OS

XX WO9506728-A.

XX

PN 09-MAR-1995.

PD

XX

PF 05-AUG-1994; 94WO-US09024.

XX

PR 13-AUG-1993; 93US-0106016.

XX

PA (IMMU-) IMMULOGIC PHARM CORP.

XX

PI Griffith IJ, Kuo M, Lugman M;

XX

XX WPI; 1995-115444/15.

DR

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s) and DNA - for treating sensitivity to ryegrass pollen allergen or an immunologically cross-reactive allergen.

PT

XX

PS Claim 1; Fig 2; 110pp; English.

XX

CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from a lambda gtlI library. Peptides (AAR71508-61) comprising at least one T cell epitope derived from the Lol pv protein are claimed, and can be used to treat or diagnose sensitivity to ryegrass pollen in an individual or to pollen proteins that are immunologically related to Lol pv, such as Dac gv (see AAR71507).

CC

XX Sequence 20 AA;

SQ

Query Match 21.9%; Score 65; DB 16; Length 20;
Best Local Similarity 60.0%; Pred. No. 0.076;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

PT or an immunologically cross-reactive allergen.

XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
CC T cell epitope derived from the Lol pv protein are claimed, and can be
CC used to treat or diagnose sensitivity to ryegrass pollen in an
CC individual or to pollen proteins that are immunologically related to Lol
CC pv, such as Dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 18.2%; Score 54; DB 16; Length 20;

Best Local Similarity 60.0%; Pred. No. 2.3;
Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 34 VLHFSEALHIIAGTPEVHAV 53

I : : : : :
Db 1 vtalearviagalevhav 20

RESULT 7

AAR89392
ID AAR89392 standard; peptide; 20 AA.

XX AAR89392;

AC AAR89392;

DT 17-SEP-1996 (first entry)

DE Kentucky Blue Grass pollen allergen rKBG60 residues 109-128.

KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
KW allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
KW allergic; allergy; prevention; development; reaction; grass;
KW antibody; anergise; producing cell; diagnosis; human; clone 60;
KW immunological response; B-cell; hlgG; hlgE; T-cell.

XX Poa pratensis.

XX WO9603106-A2.

XX 08-FEB-1996.

XX 26-JUL-1995; 95WO-CA00439.

XX 26-JUL-1994; 94US-0280455.

XX (UYMA-) UNIV MANITOBA.

XX Mohapatra SS;

XX WPI; 1996-116753/12.

PT Peptide including at least 1 human antigenic determinant of Kentucky
PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,
PT used to desensitise allergic individuals, prevent grass pollen
PT allergy and deplete allergen-specific antibodies

XX Claim 8; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide,
CC comprising residues 109-128 of the recombinant Kentucky Blue Grass
CC pollen allergen rKBG60, which is from the Poa p IX gp. of grass
CC pollen allergens. A compsn. comprising the peptide and a carrier
CC can be used to desensitise allergic individuals, and prevent
CC allergic individuals from developing an allergic reaction to grass
CC pollen. The compsn. may also be used to deplete allergen specific
CC antibodies from an individual, and anergise allergen specific
CC antibody producing cells. The peptide can be used to diagnose grass
CC pollen allergy, and has the following human immunological response

CC profile, B-cell (hlgG (-)/hlgE (-)), T-cell (+).

XX Sequence 20 AA;

Query Match 16.5%; Score 49; DB 17; Length 20;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 17 AYNADHRAAPEDKYEAFV 34

I : : : : :
Db 1 ayksaegatpeakyddv 18

RESULT 8

AAR71520
ID AAR71520 standard; Protein; 20 AA.

XX AAR71520;

AC AAR71520;

DT 02-NOV-1995 (first entry)

DE LPIX-13, peptide fragment of Lol pv protein allergen.

KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
KW sensitivity; ryegrass pollen allergen.

XX Lolium perenne.

XX WO9506728-A.

XX 09-MAR-1995.

XX 05-AUG-1994; 94WO-US09024.

XX 13-AUG-1993; 93US-0106016.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Griffith IJ, Kuo M, Luqman M;

XX WPI; 1995-115444/15.

PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
PT and DNA - for treating sensitivity to rye-grass pollen allergen
PT or an immunologically cross-reactive allergen.

XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
CC T cell epitope derived from the Lol pv protein are claimed, and can be
CC used to treat or diagnose sensitivity to ryegrass pollen in an
CC individual or to pollen proteins that are immunologically related to Lol
CC pv, such as Dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 16.2%; Score 48; DB 16; Length 20;
Best Local Similarity 83.3%; Pred. No. 14;
Matches 10; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 44 IAGTPEVHAVKP 55

I : : : : :
Db 1 iagalevhavkvp 12

RESULT 9

AAR89391
ID AAR89391 standard; peptide; 20 AA.

XX

AC	AAR89391;
XX	
DT	17-SEP-1996 (first entry)
XX	
DE	Kentucky Blue Grass pollen allergen rKBG60 residues 99-118.
XX	
KW	Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
KW	allergen; antigenic; rKBG60; P _{oa} p IX group; desensitisation;
KW	allergic; allergy; prevention; development; reaction; grass;
KW	antibody; anergise; producing cell; diagnosis; human; clone 60;
KW	immunological response; B-cell; hlgG; hlgE; T-cell.
XX	
OS	Poa pratensis.
XX	
PN	WO9603106-A2.
XX	
PD	08-FEB-1996.
XX	
PF	26-JUL-1995; 95WO-CA00439.
XX	
PR	26-JUL-1994; 94US-0280455.
XX	
PA	(UYMA-) UNIV MANITOBA.
XX	
PI	Mohapatra SS;
DR	WPI; 1996-116753/12.
XX	
PT	Peptide including at least 1 human antigenic determinant of Kentucky
PT	Blue Grass allergen - from P _{oa} p IX gp. of grass pollen allergens,
PT	used to desensitise allergic individuals, prevent grass pollen
PT	allergy and deplete allergen-specific antibodies
XX	
PS	Claim 7; Page 29; 59pp; English.
XX	
CC	The present peptide is a human antigenic determinant contg. peptide,
CC	comprising residues 99-118 of the recombinant Kentucky Blue Grass
CC	pollen allergen rKBG60, which is from the P _{oa} p IX gp. of grass
CC	pollen allergens. A compsn. comprising the peptide and a carrier
CC	can be used to desensitise allergic individuals, and prevent
CC	allergic individuals from developing an allergic reaction to grass
CC	pollen. The compsn. may also be used to deplete allergen specific
CC	antibodies from an individual, and anergise allergen specific
CC	antibody producing cells. The peptide can be used to diagnose grass
CC	pollen allergy, and has the following human immunological response
CC	profile, B-cell (hlgG (-)/hlgE (+)), T-cell (+).
XX	
SQ	Sequence 20 AA;
	Query Match 15.8%; Score 47; DB 17; Length 20;
	Best Local Similarity 50.0%; Pred. No. 19;
	Matches 9; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY	9 KLDEVYNAYNAADHAAP 26
	: :
Db	3 kldaayklaysaegatp 20
	RESULT 10
AAW76471	
ID	AAW76471 standard; peptide; 12 AA.
XX	
AC	AAW76471;
XX	
DT	11-DEC-1998 (first entry)
XX	
DE	Graminae pollen allergen Phl p 5b peptide fragment #23.
XX	
KW	Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW	epitope; immunotherapy; allergy; hyposensitisation.
XX	
OS	Graminae.

DR WPI: 2001-376931/40.
DR N-PSDB: AAH68261.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 17; SEQ ID NO: 6796; 246pp + Sequence Listing; English.
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of corynebacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Corynebacterium, and identifying a homologue of a gene derived
CC from Corynebacterium. Corynebacterium bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 53 AA;

Query Match 15.0%; Score 44.5; DB 22; Length 53;
Best Local Similarity 35.7%; Pred. No. 1.4e+02;
Matches 15; Conservative 4; Mismatches 16; Indels 7; Gaps 1;

Qy 19 NAAADHAAPEDKYE-----FVLHFSEALHIIAGTPEVHAV 53
| | | | | : : : | | | | |
Db 2 nlvdhadpensegragvmlngtlkgtdmdititagdpdvav 43

RESULT 12
AAW76473
ID AAW76473 standard; peptide; 12 AA.
XX
AC AAW76473;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #25.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
PA (MERE) MERCK PATENT GMBH.
XX
XX Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
DR WPI: 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
PT allergies
PT
PS Example 1; Page 12; 31pp; German.
XX
XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
SQ Sequence 12 AA;
Query Match 14.8%; Score 44; DB 19; Length 12;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 43 ILAGTPEVHAVK 54

CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
SQ Sequence 12 AA;
Query Match 14.8%; Score 44; DB 19; Length 12;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
Qy 7 VPKLDEVYNAAY 18
| | | | | : | |
Db 1 vpkldaaysvay 12

RESULT 13
AAW76485
ID AAW76485 standard; peptide; 12 AA.
XX
AC AAW76485;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #37.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
OS Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
PA (MERE) MERCK PATENT GMBH.
XX
XX Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
DR WPI: 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
PT allergies
PT
PS Example 1; Page 12; 31pp; German.
XX
XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.
XX
SQ Sequence 12 AA;

Query Match 14.8%; Score 44; DB 19; Length 12;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy 43 ILAGTPEVHAVK 54

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2002, 10:50:47 ; Search time 29.69 seconds

(without alignments)

299.289 Million cell updates/sec

Title: US-09-696-169A-18

Perfect score: 409

Sequence: 1 ADKYTFEAFATVSKRNLA.....SEALHIIAGTPEVHAVKPGA 80

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 434996

Minimum DB seq length: 0

Maximum DB seq length: 80

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	409	100.0	80	AA125638	Phleum sp. allerg
2	297	72.6	57	AA125637	Phleum sp. allerg
3	70	17.1	20	AA189393	Kentucky Blue Gras
4	65	15.9	20	AA171517	LPIX-10, peptide f
5	62	15.2	20	AA171516	LPIX-9, peptide fr
6	59	14.4	20	AA171513	LPIX-6, peptide fr
7	56	13.7	20	AA171518	LPIX-11, peptide f
8	54	13.2	12	AA176465	Graminae pollen al
9	54	13.2	20	AA171519	LPIX-12, peptide f
10	52	12.7	20	AA171515	LPIX-8, peptide fr
11	51	12.5	12	AA176466	Graminae pollen al

12	50.5	12.3	76	22	ABG06223	Novel human diagno
13	49.5	12.1	76	21	AA10561	Copolymer block pr
14	49	12.0	20	17	AA189392	Kentucky Blue Gras
15	48.5	11.9	56	21	AA182573	Copolymer molecula
16	48.5	11.9	69	22	AAU46409	Propionibacterium
17	48.5	11.9	76	21	AA10562	Copolymer block pr
18	48	11.7	20	16	AA171520	LPIX-13, peptide f
19	47	11.5	12	19	AA176464	Graminae pollen al
20	47	11.5	20	17	AA189391	Kentucky Blue Gras
21	47	11.5	59	21	AA189391	Human secreted pro
22	47	11.5	67	22	AA189391	Novel human diagno
23	47	11.5	78	21	AA189391	Human secreted pro
24	47	11.5	78	21	AA189391	Human secreted pro
25	46.5	11.4	66	21	AA182574	Copolymer molecula
26	46	11.2	20	17	AA189391	Kentucky Blue Gras
27	46	11.2	67	20	AA111961	Human 5' EST secre
28	46	11.2	68	22	AA162884	Alpha-tocopherol t
29	46	11.2	73	22	AAU53006	Propionibacterium
30	46	11.2	77	22	AB166772	Drosophila melanog
31	45.5	11.1	51	21	AA102471	Human secreted pro
32	45.5	11.1	59	22	AAU64369	Propionibacterium
33	45.5	11.1	61	20	AA110914	Amino acid sequenc
34	45.5	11.1	69	21	AA12165	Arabidopsis thalia
35	45.5	11.1	69	21	AA120982	Arabidopsis thalia
36	45.5	11.1	69	21	AA146642	Arabidopsis thalia
37	45	11.0	12	19	AA176471	Graminae pollen al
38	45	11.0	56	22	AB132073	Peptide #4724 enco
39	45	11.0	56	22	AB137325	Peptide #4831 enco
40	45	11.0	56	22	AB122615	Protein #4614 enco
41	45	11.0	56	22	AA158018	Human brain expres
42	45	11.0	56	22	AA170455	Human bone marrow
43	45	11.0	56	22	AA118290	Peptide #4724 enco
44	45	11.0	56	22	AA130778	Peptide #4815 enco
45	45	11.0	56	22	AA105901	Peptide #4583 enco

ALIGNMENTS

RESULT 1

AA125638
ID AA125638 standard; protein; 80 AA.

XX AC AA125638;

XX AC AA125638;

DT 30-SEP-1999 (first entry)

XX Phleum sp. allergen Phl p 6 protein fragment #3.

DE Major histocompatibility complex; class II; desensitising; human;
XX allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting;
KW chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee;
KW screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat;
KW cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig;
KW mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Phleum sp.

XX WO9934826-A1.

PD 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

DR WPI; 1999-458255/38.

PT Desensitizing patients to polypeptide allergens

PS Example 6; Page 65-66; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tenbriol molitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a Phleum sp. (Timothy grass) Phl p 6 allergen.

XX Sequence 80 AA;

Query Match 100.0%; Score 409; DB 20; Length 80;

Best Local Similarity 100.0%; Pred. No. 1.3e-42;

Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ADKYKTFEAFVSSKRNLADAVSKAPQLVPLKDEVYNAAYNAADHAAPEDKYEAFVLHF 60

Db 1 adkyktfeafvsskrnladvskapqlvplkdevynaaynaadhaapedkyeatvlhf 60

QY 61 SEALHIIAGTPEVHAVKPGA 80

Db 61 sealhiiagtpevhavkpga 80

RESULT 2

AAY25637

ID AAY25637 standard; protein; 57 AA.

AC AAY25637;

DT 30-SEP-1999 (first entry)

DE Phleum sp. allergen Phl p 6 protein fragment #2.

XX Major histocompatibility complex; class II; desensitising; human; allergen; grass; tree; weed; pollen; fungi; mould; food; insect; sting; chironomidae; spider; mite; housefly; fruit fly; sheep blow fly; honeybee; screw worm fly; grain weevil; silkworm; bee moth; larvae; mealworm; cat; cockroach; beetle; dog; horse; cow; pig; sheep; rabbit; rat; guinea pig; mice; gerbil; vaccine; treatment; prevention; hypersensitivity.

XX Phleum sp.

OS WO9934826-A1.

PN 15-JUL-1999.

XX 11-JAN-1999; 99WO-GB00080.

XX 21-SEP-1998; 98GB-0020474.

PR 09-JAN-1998; 98GB-0000445.

XX (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.

XX Kay AB, Larche M;

PI WPI; 1999-458255/38.

XX Desensitizing patients to polypeptide allergens

XX

PS Example 6; Page 65; 117pp; English.

XX This invention describes a novel method of desensitizing a patient to a polypeptide allergen and comprises administering to the patient a peptide derived from the allergen where restriction to a MHC Class II molecule possessed by the patient can be demonstrated for the peptide and the peptide is able to induce a late phase response in an individual who possesses the MHC Class II molecule. The methods can be used for desensitizing patients to allergens present in e.g. grass, tree and weed (including ragweed) pollens, fungi and moulds, foods, stinging insects, the chironomidae (non-biting midges), spiders and mites, housefly, fruit fly, sheep blow fly, screw worm fly, grain weevil, silkworm, honeybee, non-biting midge larvae, bee moth larvae, mealworm, cockroach, larvae of tenbriol molitor beetle, mammals such as cat, dog, horse, cow, pig, sheep, rabbit, rat, guinea pig, mice or gerbil. They can also be used to produce immunological vaccines which may be used to prevent and/or treat conditions involving hypersensitivity to allergens. This sequence represents a Phleum sp. (Timothy grass) Phl p 6 allergen.

XX Sequence 57 AA;

Query Match 72.6%; Score 297; DB 20; Length 57;

Best Local Similarity 100.0%; Pred. No. 4.6e-29;

Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 SKAPOLVPKLDEVYNAAYNAADHAAPEDKYEAFVLHFSEALHIIAGTPEVHAVKPGA 80

Db 1 skapqlvplkdevynaaynaadhaapedkyeatvlhfsealhiagtpevhavkpga 57

RESULT 3

AAR89393

ID AAR89393 standard; peptide; 20 AA.

XX AAR89393;

XX 17-SEP-1996 (first entry)

DT Kentucky Blue Grass pollen allergen rKBG60 residues 129-148.

DE Antigenic determinant; recombinant; Kentucky Blue Grass; pollen; allergen; antigenic; rKBG60; Poa p IX group; desensitisation; KW allergic; allergy; prevention; development; reaction; grass; KW antibody; anergise; producing cell; diagnosis; human; clone 60; KW immunological response; B-cell; hlg; hlgE; T-cell.

XX Poa pratensis.

OS WO9603106-A2.

XX 08-FEB-1996.

XX 26-JUL-1995; 95WO-CA00439.

XX 26-JUL-1994; 94US-0280455.

XX (UYMA-) UNIV MANITOBA.

XX Mohapatra SS;

PI WPI; 1996-116753/12.

XX Peptide including at least 1 human antigenic determinant of Kentucky

PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens, used to desensitise allergic individuals, prevent grass pollen allergy and deplete allergen-specific antibodies

XX Example 5; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide, comprising residues 129-148 of the recombinant Kentucky Blue Grass pollen allergen rKBG60, which is from the Poa p IX gp. of grass

CC pollen allergens. A compasn. comprising the peptide and a carrier
 CC can be used to desensitize allergic individuals, and prevent
 CC allergic individuals from developing an allergic reaction to grass
 CC pollen. The compsn. may also be used to deplete allergen specific
 CC antibodies from an individual, and anergise allergen specific
 CC antibody producing cells. The peptide can be used to diagnose grass
 CC pollen allergy, and has the following human immunological response
 CC profile, B-cell (hlgG (-)/hlgE (-)), T-cell (-).
 XX
 SQ Sequence 20 AA;

Query Match 17.1%; Score 70; DB 17; Length 20;
 Best Local Similarity 83.3%; Pred. No. 0.085;
 Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 61 SEALHIIAGTPEVHAVKP 78
 DB 2 sealriiagtlevehgvp 19
 ||||| ||||| |||||

RESULT 4
 AAR71517
 ID AAR71517 standard; Protein; 20 AA.
 AC
 AAR71517;
 XX
 DT 02-NOV-1995 (first entry)
 XX
 DE LPIX-10, peptide fragment of Lol pv protein allergen.
 XX
 KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 KW sensitivity; ryegrass pollen allergen.
 XX
 OS Lolium perenne.
 XX
 PN WO9506728-A.
 XX
 PD 09-MAR-1995.
 XX
 PF 05-AUG-1994; 94WO-US09024.
 XX
 PR 13-AUG-1993; 93US-0106016.
 XX
 PA (IMMU-) IMMULOGIC PHARM CORP.
 XX
 PI Griffith IJ, Kuo M, Lugman M;
 XX
 PS WPI; 1995-115444/15.
 XX
 PD 09-MAR-1995.
 XX
 PF 05-AUG-1994; 94WO-US09024.
 XX
 PR 13-AUG-1993; 93US-0106016.
 XX
 PA (IMMU-) IMMULOGIC PHARM CORP.
 XX
 PI Griffith IJ, Kuo M, Lugman M;
 XX
 PS WPI; 1995-115444/15.
 XX
 PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 PT and DNA - for treating sensitivity to rye-grass pollen allergen
 PT or an immunologically cross-reactive allergen.
 XX
 PS Claim 1; Fig 2; 110pp; English.
 XX
 CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).
 XX
 SQ Sequence 20 AA;

Query Match 15.9%; Score 65; DB 16; Length 20;
 Best Local Similarity 60.0%; Pred. No. 0.35;
 Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 37 YNAAYNAADHAAPEDKYEF 56
 |: || ||: | || ||: ||

DB 1 ydvaykaeagatpeakydaf 20

RESULT 5
 AAR71516
 ID AAR71516 standard; Protein; 20 AA.
 XX
 AC AAR71516;
 XX
 DT 02-NOV-1995 (first entry)
 XX
 DE LPIX-9, peptide fragment of Lol pv protein allergen.
 XX
 KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 KW sensitivity; ryegrass pollen allergen.
 XX
 OS Lolium perenne.
 XX
 PN WO9506728-A.
 XX
 PD 09-MAR-1995.
 XX
 PF 05-AUG-1994; 94WO-US09024.
 XX
 PR 13-AUG-1993; 93US-0106016.
 XX
 PA (IMMU-) IMMULOGIC PHARM CORP.
 XX
 PI Griffith IJ, Kuo M, Lugman M;
 XX
 PS WPI; 1995-115444/15.
 XX
 PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 PT and DNA - for treating sensitivity to rye-grass pollen allergen
 PT or an immunologically cross-reactive allergen.
 XX
 PS Claim 1; Fig 2; 110pp; English.
 XX
 CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 CC cDNA sequence of clone 12R (AAQ85932), a full-length clone derived from
 CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 CC T cell epitope derived from the Lol pv protein are claimed, and can be
 CC used to treat or diagnose sensitivity to ryegrass pollen in an
 CC individual or to pollen proteins that are immunologically related to Lol
 CC pv, such as Dac gv (see AAR71507).
 XX
 SQ Sequence 20 AA;

Query Match 15.2%; Score 62; DB 16; Length 20;
 Best Local Similarity 57.9%; Pred. No. 0.81;
 Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 27 POLVPKLDEVYNAAYNAAD 45
 |: ||||| |: |||:
 DB 1 pglipkldtaydvaykaae 19

RESULT 6
 AAR71513
 ID AAR71513 standard; Protein; 20 AA.
 XX
 AC AAR71513;
 XX
 DT 02-NOV-1995 (first entry)
 XX
 DE LPIX-6, peptide fragment of Lol pv protein allergen.
 XX
 KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 KW sensitivity; ryegrass pollen allergen.
 XX
 OS Lolium perenne.


```

PN WO9506728-A.
XX
PD 09-MAR-1995.
XX
PF 05-AUG-1994; 94WO-US09024.
XX
PR 13-AUG-1993; 93US-0106016.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Griffith IJ, Kuo M, Lugman M;
XX
DR WPI; 1995-115444/15.
XX
PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
PT and DNA - for treating sensitivity to rye-grass pollen allergen
PT or an immunologically cross-reactive allergen.
XX
PS Claim 1; Fig 2; 110pp; English.
XX
CC Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
CC cDNA sequence of clone 12R (AA085932), a full-length clone derived from
CC a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
CC T cell epitope derived from the Lol pv protein are claimed, and can be
CC used to treat or diagnose sensitivity to ryegrass pollen in an
CC individual or to pollen proteins that are immunologically related to Lol
CC pv, such as Dac gv (see AAR71507).
XX
XX
XX Sequence 20 AA;
XX
Query Match 14.4%; Score 59; DB 16; Length 20;
Best Local Similarity 75.0%; Pred. No. 1.9;
Matches 12; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 ADKYKTFEAAFTVSSK 16
DB 5 adkfkieafesssk 20
|||||
| | | | |

RESULT 7
AAR71518
ID AAR71518 standard; Protein; 20 AA.
XX
AC AAR71518;
XX
DT 02-NOV-1995 (first entry)
XX
DE LP1X-11, peptide fragment of Lol pv protein allergen.
XX
KW Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
KW sensitivity; ryegrass pollen allergen.
XX
OS Lolium perenne.
XX
XX WO9506728-A.
XX
XX 09-MAR-1995.
XX
PF 05-AUG-1994; 94WO-US09024.
XX
PR 13-AUG-1993; 93US-0106016.
XX
PA (IMMU-) IMMULOGIC PHARM CORP.
XX
PI Griffith IJ, Kuo M, Lugman M;
XX
DR WPI; 1995-115444/15.
XX
PT Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
PT and DNA - for treating sensitivity to rye-grass pollen allergen
PT or an immunologically cross-reactive allergen.
XX
XX

```

```

PS Claim 1; Fig 2; 110pp; English.
XX
XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
XX cDNA sequence of clone 12R (AA085932), a full-length clone derived from
XX a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
XX T cell epitope derived from the Lol pv protein are claimed, and can be
XX used to treat or diagnose sensitivity to ryegrass pollen in an
XX individual or to pollen proteins that are immunologically related to Lol
XX pv, such as Dac gv (see AAR71507).
XX
XX Sequence 20 AA;
XX
Query Match 13.7%; Score 56; DB 16; Length 20;
Best Local Similarity 55.0%; Pred. No. 4.4;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 47 AAPEDKYEARVLFHFSEALHI 66
DB 1 atpekydafvtalr 20
|||||
| | | | |

RESULT 8
AAW76465
ID AAW76465 standard; peptide; 12 AA.
XX
XX AC AAW76465;
XX
DT 11-DEC-1998 (first entry)
XX
DE Graminae pollen allergen Phl p 5b peptide fragment #17.
XX
KW Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
KW epitope; immunotherapy; allergy; hyposensitisation.
XX
XX Graminae.
XX
PN DE19713001-A1.
XX
PD 01-OCT-1998.
XX
PF 27-MAR-1997; 97DE-1013001.
XX
PR 27-MAR-1997; 97DE-1013001.
XX
PA (MERE ) MERCK PATENT GMBH.
XX
PI Becker W, Bufer A, Cromwell O, Fiebig H, Jaeger L;
PI Kahlert H, Mueller W, Schramm G, Stuewe H;
XX
DR WPI; 1998-522170/45.
XX
XX Modified recombinant allergens - useful for immuno-therapy of
XX allergies
XX
XX Example 1; Page 12; 31pp; German.
XX
XX AAW76449-W76534 are peptide fragments of a natural pollen allergen,
XX Phl p 5b, isolated from graminiae species. This allergen can be modified
XX and the reactivity of the modified allergens with IgE antibodies to
XX grass pollen allergens is reduced or eliminated while their reactivity
XX with T cells is retained. The genes for the allergens are modified so
XX that the encoded polypeptides have one or more amino acid substitutions,
XX deletions and/or additions. The dominant T-cell epitopes of the
XX allergens are not genetically altered. Such allergens have applications
XX in the immunotherapy of allergies e.g. hyposensitisation.
XX
XX Sequence 12 AA;
XX
Query Match 13.2%; Score 54; DB 19; Length 12;
Best Local Similarity 90.9%; Pred. No. 4;
Matches 10; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 DKYKTEFAFT 12
 ||:|||||
 Db 1 dkfkfefaft 11

RESULT 9

AAR71519
 ID AAR71519 standard; Protein; 20 AA.

XX AAR71519;

AC AAR71519;

DT 02-NOV-1995 (first entry)

DE LP1X-12, peptide fragment of Lol pv protein allergen.

XX Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 sensitivity; ryegrass pollen allergen.

XX Lolium perenne.

PN WO9506728-A.

XX 09-MAR-1995.

PF 05-AUG-1994; 94WO-US09024.

XX 13-AUG-1993; 93US-0106016.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Griffith LJ, Kuo M, Luqman M;

XX WPI; 1995-115444/15.

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 and DNA - for treating sensitivity to rye-grass pollen allergen
 or an immunologically cross-reactive allergen.

XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 cDNA sequence of clone 12R (AAO85932), a full-length clone derived from
 a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 T cell epitope derived from the Lol pv protein are claimed, and can be
 used to treat or diagnose sensitivity to ryegrass pollen in an
 individual or to pollen proteins that are immunologically related to Lol
 pv, such as Dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 13.2%; Score 54; DB 16; Length 20;

Best Local Similarity 60.0%; Pred. No. 7.8;

Matches 12; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 57 VLHFEALHIIAGTPPEVHAV 76

Db 1 vtalearlrviaalevhav 20

RESULT 10

AAR71515

ID AAR71515 standard; Protein; 20 AA.

XX AAR71515;

XX 02-NOV-1995 (first entry)

DE LP1X-8, peptide fragment of Lol pv protein allergen.

XX Lolium perenne; Lol pv; Dactylis glomerata; Dac gv; epitope;
 sensitivity; ryegrass pollen allergen.

XX

OS Lolium perenne.

XX WO9506728-A.

PN 09-MAR-1995.

PD 05-AUG-1994; 94WO-US09024.

XX 13-AUG-1993; 93US-0106016.

XX (IMMU-) IMMULOGIC PHARM CORP.

XX Griffith LJ, Kuo M, Luqman M;

XX WPI; 1995-115444/15.

XX Lolium perenne Lol pv and Dactylis glomerata Dac gv epitope(s)
 and DNA - for treating sensitivity to rye-grass pollen allergen
 or an immunologically cross-reactive allergen.

XX Claim 1; Fig 2; 110pp; English.

XX Lol pv, is a major allergen of ryegrass pollen, and is encoded by the
 cDNA sequence of clone 12R (AAO85932), a full-length clone derived from
 a lambda gt11 library. Peptides (AAR71508-61) comprising at least one
 T cell epitope derived from the Lol pv protein are claimed, and can be
 used to treat or diagnose sensitivity to ryegrass pollen in an
 individual or to pollen proteins that are immunologically related to Lol
 pv, such as Dac gv (see AAR71507).

XX Sequence 20 AA;

Query Match 12.7%; Score 52; DB 16; Length 20;

Best Local Similarity 62.5%; Pred. No. 14;

Matches 10; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 19 LADAVSKAPQLVPKLD 34

Db 3 latsaakapglipkld 18

XX

RESULT 11

AAW76466

ID AAW76466 standard; peptide; 12 AA.

XX AAW76466;

XX 11-DEC-1998 (first entry)

XX Graminae pollen allergen Phl p 5b peptide fragment #18.

XX Pollen; allergen; Phl p 5b; IgE; antibody; grass; reactivity; T cell;
 epitope; immunotherapy; allergy; hyposensitisation.

XX Graminae.

XX DE19713001-A1.

XX 01-OCT-1998.

XX 27-MAR-1997; 97DE-1013001.

XX 27-MAR-1997; 97DE-1013001.

XX (MERE) MERCK PATENT GMBH.

XX Becker W, Bufo A, Cromwell O, Fiebig H, Jaeger L;

XX Kahlert H, Mueller W, Schramm G, Stuewe H;

XX WPI; 1998-522170/45.

XX

PT Modified recombinant allergens - useful for immuno-therapy of
PT allergies

XX Example 1; Page 12; 31pp; German.

XX AA076449-W76534 are peptide fragments of a natural pollen allergen,
CC Phl p 5b, isolated from Graminae species. This allergen can be modified
CC and the reactivity of the modified allergens with IgE antibodies to
CC grass pollen allergens is reduced or eliminated while their reactivity
CC with T cells is retained. The genes for the allergens are modified so
CC that the encoded polypeptides have one or more amino acid substitutions,
CC deletions and/or additions. The dominant T-cell epitopes of the
CC allergens are not genetically altered. Such allergens have applications
CC in the immunotherapy of allergies e.g. hyposensitisation.

XX Sequence 12 AA;

Query Match 12.5%; Score 51; DB 19; Length 12;
Best Local Similarity 91.7%; Pred. No. 9.4;
Matches 11; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KTFEAAFTVSSK 16

DB 1 ktfeaaftssk 12

RESULT 12

ABG06223

ID ABG06223 standard; Protein; 76 AA.

AC ABG06223;

DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #6214.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS70410.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX Claim 20; SEQ ID No 36582; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 76 AA;

Query Match 12.3%; Score 50.5; DB 22; Length 76;
Best Local Similarity 34.2%; Pred. No. 1.1e+02;
Matches 13; Conservative 6; Mismatches 16; Indels 3; Gaps 1;

QY 23 VSKAP--QLVPKLDEVYNAAYNAADHAAPEDKYEAFV 57

DB 13 vkskpvkhmcvevldsvyssgpnrdhpxtsvowelym 50

RESULT 13

AAB10561

ID AAB10561 standard; protein; 76 AA.

AC AAB10561;

DT 15-DEC-2000 (first entry)

XX Copolymer block protein insert L2-A.

XX Coiled coil protein; copolymer blocker; vulnery; wound treatment;
KW integrin; fibroblast cell; matrix formation; skin regeneration; scaffold;
KW antibiotic; dressing; abrasion; burn.

OS Unidentified.

XX US6090911-A.

XX 18-JUL-2000.

XX 22-OCT-1997; 97US-0956307.

XX 22-OCT-1997; 97US-0956307.

XX (UYMA-) UNIV MASSACHUSETTS.

XX Petka WA, McGrath KP, Tirrell DA;

XX WPI; 2000-542235/49.

XX Synthetic block copolymer with an antibiotic compound, useful for
PT dressing abrasion, burn or non-puncture wound, comprises two
PT alpha-helical protein blocks, random-coil protein block and linker
PT proteins -

XX Example 1; Column 33-34; 53pp; English.

XX This invention describes a novel synthetic block copolymer XYZ comprising
CC two alpha-helical protein blocks X and Z, a water soluble, random-coil
CC protein block Y which links X and Z, and linker proteins for linking the
CC alpha-helical protein blocks to the random-coil protein block. The
CC products of the invention have vulnery activity. A copolymer produced
CC from a gene that encodes a derivative of AC1.0A copolymer was suspended
CC in water to form a gel, which was then used to treat a wound. The cell
CC binding domain was the integrin ArgGlyAspser sequence which is known to
CC bind to gp130/131a proteins expressed on fibroblast cells necessary for
CC matrix formation for the regeneration of skin. Fibroblasts become

CC entrapped within the gel and thus remain at the site of the wound,
CC serving as a scaffold for the regeneration of tissue surrounding the
CC wound. A synthetic block copolymer together with an antibiotic compound
CC is useful for wound dressing. The copolymer is useful for dressing
CC abrasion, burn or non-puncture wound. Unlike conventional polymers, the
CC polymers of equal size are produced biologically from a single template.
CC The intermolecular binding of alpha-helical blocks are monodisperse which
CC gives a uniform pore size depending on the length of random-coil block.
CC The new copolymers have low molecular weight hence have decreased
CC viscosity solutions or gels under suitable conditions. This sequence
CC represents a copolymer block protein L2-A which is used to illustrate the
CC method of the invention.
XX
SQ Sequence 76 AA;

Query Match 12.1%; Score 49.5; DB 21; Length 76;
Best Local Similarity 33.3%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 16; Indels 1; Gaps 1;

QY 17 RNLADAVSKAPOLYVKL-DEVYNAAYNAADHAAPED 51
| | | | : : | | : | : : | | | |
Db 35 rsldeaaealeqkvsrlkneiedikaeighdvaprd 70

RESULT 14

AAR89392
ID AAR89392 standard; peptide; 20 AA.

AC AAR89392;

DT 17-SEP-1996 (first entry)

DE Kentucky Blue Grass pollen allergen rKBG60 residues 109-128.

KW Antigenic determinant; recombinant; Kentucky Blue Grass; pollen;
KW allergen; antigenic; rKBG60; Poa p IX group; desensitisation;
KW allergic; allergy; prevention; development; reaction; grass;
KW antibody; anergise; producing cell; diagnosis; human; clone 60;
KW immunological response; B-cell; hlgG; hlgE; T-cell.

OS Poa pratensis.

PN WO9603106-A2.

PD 08-FEB-1996.

PF 26-JUL-1995; 95WO-CA00439.

PR 26-JUL-1994; 94US-0280455.

PA (UYMA-) UNIV MANITOBA.

PI Mohapatra SS;

DR WPI; 1996-116753/12.

PT Peptide including at least 1 human antigenic determinant of Kentucky
PT Blue Grass allergen - from Poa p IX gp. of grass pollen allergens,
PT used to desensitise allergic individuals, prevent grass pollen
PT allergy and deplete allergen-specific antibodies

PS Claim 8; Page 29; 59pp; English.

XX The present peptide is a human antigenic determinant contg. peptide,
CC comprising residues 109-128 of the recombinant Kentucky Blue Grass
CC pollen allergen rKBG60, which is from the Poa p IX gp. of grass
CC pollen allergens. A compsn. comprising the peptide and a carrier
CC can be used to desensitise allergic individuals, and prevent
CC allergic individuals from developing an allergic reaction to grass
CC pollen. The compsn. may also be used to deplete allergen specific
CC antibodies from an individual, and anergise allergen specific
CC antibody producing cells. The peptide can be used to diagnose grass

CC pollen allergy, and has the following human immunological response
CC profile, B-cell (hlgG (-)/hlgE (-)), T-cell (+).
XX
SQ Sequence 20 AA;

Query Match 12.0%; Score 49; DB 17; Length 20;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 9; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 40 AYNAADHAAPEDKYAEFV 57

| | : | : | | | : |

Db 1 ayksaegatpeakyddv 18

RESULT 15

AAY82573

ID AAY82573 standard; peptide; 56 AA.

AC AAY82573;

DT 28-JUL-2000 (first entry)

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.

KW Copolymer; molecular weight marker; TV-marker; immune disease;
KW glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW antiinflammatory; immunosuppressive; demyelinating disease; osteoarthritis;
KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;
KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW pemphigus vulgaris; systemic lupus erythematosus.

OS Unidentified.

PN WO200018794-A1.

PD 06-APR-2000.

PF 24-SEP-1999; 99WO-US22402.

PR 25-SEP-1998; 98US-0101693.

PA (YEDA) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.

PI Gad A, Lis D;

WPI; 2000-317499/27.

PT Copolymer 1 related polypeptides used as molecular weight markers for
PT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.

XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glatiramer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia
CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves

CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glatiramer acetate molecules, which makes them ideal for use as
XX molecular weight markers.

SQ Sequence 56 AA;

Query Match 11.9%; Score 48.5; DB 21; Length 56;
Best Local Similarity. 32.8%; Pred. No. 1.4e+02;
Matches 19; Conservative 7; Mismatches 27; Indels 5; Gaps 2;

QY 1 ADKYKTFEAAFTVSSKRNLADAVS---KAPQLVPKLDEVYNNAYNAADHAAPEDKYEA 55
||| | | : : : | : | | | | | | |
Db 1 akkyakkekayakkaekakkaekaykaeakkkkaekyka--eaakaakeaayea 56

Search completed: July 11, 2002, 10:51:26
Job time: 39 sec